

Article

A Hybrid Method with TOPSIS and Machine Learning Techniques for Sustainable Development of Green Hotels Considering Online Reviews

Mehrbakhsh Nilashi ^{1,2}, Abbas Mardani ³, Huchang Liao ^{4,5,*}, Hossein Ahmadi ⁶, Azizah Abdul Manaf ⁷ and Wafa Almukadi ⁸

- ¹ Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City 758307, Vietnam; nilashi@tdtu.edu.vn
 - ² Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City 758307, Vietnam
 - ³ Department of Marketing, College of Business Administration, University of South Florida, Tampa, FL 33813, USA; mabbas3@liveutm.onmicrosoft.com
 - ⁴ Business School, Sichuan University, Chengdu 610064, China
 - ⁵ Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 Granada, Spain
 - ⁶ Department of Information Technology, University of Human Development, Sulaymaniyah 00964, Iraq; hossein.ahmadi@uhd.edu.iq
 - ⁷ Department of Cybersecurity, College of Computer Science and Engineering, University of Jeddah, Jeddah 23218, Saudi Arabia; aaabdmanaf@uj.edu.sa
 - ⁸ Department of Software Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 23218, Saudi Arabia; wsalmukadi@uj.edu.sa
- * Correspondence: liaohuchang@scu.edu.cn or liaohuchang@163.com; Tel.: +86-177-7161-1031

Received: 28 August 2019; Accepted: 25 October 2019; Published: 29 October 2019



Abstract: This paper proposes a hybrid method for online reviews analysis through multi-criteria decision-making, text mining and predictive learning techniques to find the relative importance of factors affecting travelers' decision-making in selecting green hotels with spa services. The proposed method is developed for the first time in the context of tourism and hospitality by this research, especially for customer segmentation in green hotels through customers' online reviews. We use Self-Organizing Map (SOM) for cluster analysis, Latent Dirichlet Analysis (LDA) technique for analyzing textual reviews, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) for ranking hotel features, and Neuro-Fuzzy technique to reveal the customer satisfaction levels. The impact of green hotels with spa and non-spa services on travelers' satisfaction is investigated for four travelling groups: Travelled solo, Travelled with family, Travelled as a couple and Travelled with friends. The proposed method is evaluated on the travelers' reviews on 152 hotels in Malaysia. The findings of this study provide an important method for travelers' decision-making for hotel selection through User-Generated Content (UGC) and help hotel managers to improve their service quality and marketing strategies.

Keywords: sustainable development; green hotels; multi-criteria decision-making; TOPSIS; machine learning techniques; neuro-fuzzy; big data; satisfaction

1. Introduction

Over the past decade, green practices have been important in the hospitality and tourism sector [1,2] and gained much market interest in this sector for green or eco-friendly products/services [3,4]. Because of the increased interest among customers in tourism, friendly products/services have attracted much

attraction among hotel managers and service providers [5]. To be successful in this marketing, detecting the customer behavioral intentions in purchasing green products is important as it can assist them in the decision-making process to receive high-quality services [6]. In addition, it is important to investigate the factors for tourists' intentions to participate in green tourism [7].

Green tourism is defined as "tourism that enhances the local cultural elements, operating under the control of local communities, providing employment and maintaining the economic benefits within the local communities" [8]. Green tourism and its associated businesses have been one of the most profitable hospitality sectors for many tourist destinations. In such a progressively competitive market, the main emphasis of hoteliers and practitioners is to magnetize more green travelers (i.e., travelers who are willing to stay in environmentally friendly hotels) by motivating them to make repeat purchases through service strategies/efforts. Therefore, identifying important factors influencing green tourists' purchase decision-making behaviour and exploring their specific role are becoming more and more critical for any destination country and its related green economy sectors.

Customers' satisfaction has long been regarded as a significant factor in explaining customer purchase behavior. This variable actively contributes to forming favorable intentions toward a business and affecting loyalty and retention of customers [9]. If users assess their overall consumption experiences positively, their level of satisfaction and readiness/willingness to repurchase will rise [10]. Providing satisfactory experiences for tourists is a key success in the international tourism market, as well as in the wellness tourism industry. Studies in the wellness tourism market consider satisfaction as a cognitive or emotional judgment prior to purchase [11,12] and after consumption evaluation [13]. However, in marketing literature, satisfaction is defined as "a favorable reaction emerging from a positive assessment of consumption experiences" [14]. This definition is more commonly agreed upon by many researchers, and thus generally followed in existing literature [15]. Despite the importance of customer satisfaction, studies surrounding travellers' behavior intentions remains insufficient in the literature [16]. In spite of the significant progress in the tourism market, only a limited number of studies focused on users' behaviour, preference, and satisfaction toward green tourism products. In fact, a question still remains with regard to the causal relationship between the green hotels with wellness services and customer satisfaction.

The process of extracting consensus opinions from online reviews and analyzing big amounts of travellers' information in social networking sites has been a challenge in many studies in the tourism and hospitality context. In fact, extracting the dimensions of satisfaction through opinion mining from text-based review comments has been one of the main aims of these studies [17,18]. This study continues the previous work on the online reviews analysis for extracting useful information from the past users' experiences on the quality of green hotels services and their facilities for understanding the consumer preferences and demand. We focus on the online reviews and ratings of travellers on the green hotels with spa services and non-spa hotels. We further investigate the role of eco-friendly (green) products on the travellers' satisfaction. In addition, we try to reveal the impact of eco-friendly services on four types of travelers: Travelled solo, Travelled with family, Travelled as a couple and Travelled with friends. From a marketing perspective, the proposed method can be effective in customers' segmentation to assist the hotels' managers in better understanding the customers' need in eco-friendly hotels.

To perform the data analysis, we apply several data mining and learning techniques. To better extract the knowledge from the travellers' reviews, we adopt a clustering analysis. In addition, to effectively extract the dimensions of satisfaction from the hotels' reviews, we use an unsupervised learning approach, Latent Dirichlet Analysis (LDA), which is based on Bayesian learning algorithm. To find the importance level of spa service on the customers' satisfaction, we use Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) for ranking hotel features and a neuro-fuzzy algorithm, Adaptive Neuro-Fuzzy Inference System (ANFIS) in Matlab (R2016a, the Mathworks Inc., Natic, MA, USA), to learn from the extracted dimensions and general attributes of the eco-friendly hotels with spa and non-spa services. Therefore, the research hypothesis which this research aims

to investigate is: “There is a difference in the customers’ satisfaction level in Malaysian eco-friendly hotels with spa and non-spa services” Overall, the contributions of this work are as follows:

We develop a new method through clustering, text mining, multi-criteria decision-making and supervised learning approaches for customer segmentation in eco-friendly hotels. Multi-criteria decision-making approaches are effective in assessing the alternatives where many criteria are available. The combination of these techniques for the customer segmentation in eco-friendly hotels is developed for the first time in the context of tourism and hospitality by this research. In contrast with previous research which is mainly based on qualitative and quantitative statistical analysis through the interview and survey, the proposed method uses machine learning techniques for revealing the customers’ satisfaction level and predicting their preferences through customers’ online reviews.

We use the clustering approach to segment the customers according to their online reviews on the quality aspect of green hotels. In addition, we use Self-Organizing Map (SOM) [19], which is a robust clustering technique based on the neural network approach for customers’ segmentation from big datasets.

- i. We use a text mining approach, LDA [20], to discover satisfaction dimensions from text-based online reviews. The LDA has shown its effectiveness in text-based reviews in e-commerce [21] and tourism and hospitality [17]. Detecting customers’ behaviour from user-generated content is important as, in many social networking sites, this way is extensively used to collect customers’ feedback on the products. In the context of non-green tourism, several studies have been conducted to reveal the customers’ satisfaction from customers’ online reviews [22,23]. However, there is a lack of research on the detection of customers’ behaviour in green tourism from online reviews. This research fills this gap by developing a new method using the LDA with the aid of supervised machine learning technique.
- ii. We use a Multi-Criteria Decision-Making approach, TOPSIS [24], for ranking hotel features in each segment discovered by SOM clustering technique. This technique has been widely used in decision-making problems [24–26]. In contrast with the other Multi-Criteria Decision-Making approach (e.g., Analytic Hierarchy Process and Analytic Network Process), which are based on pairwise comparisons, TOPSIS technique is suitable for data analysis of this research as the data are collected in a five-point Likert scale to represent the travelers’ preferences.
- iii. We use a neuro-fuzzy approach, ANFIS [27], to predict the customers’ preferences through their online reviews in each segment discovered by SOM clustering technique. In this research, ANFIS, which takes the advantages of neural network and fuzzy logic approaches to learn the prediction models, is used for preference prediction through customers’ numerical ratings.

The rest of the paper is organized as follows. Section 2 provides related work. Section 3 provides the proposed research method. In Section 4, we provide the results of our data analysis. In Section 5, we provide the discussions. Finally, discussions and conclusions are provided in Section 6.

2. Related Work

Customer segmentation has been important in marketing research [28,29]. Accordingly, many methods have been developed by machine learning, statistical, and decision-making techniques in this field [30–32]. Such methods have also been effective in the tourism context, especially in sustainable and wellness tourism [33].

In [32], the authors focused on the customer knowledge mining for tourism for new product development. They used the Apriori algorithm to discover association rules and clustering analysis for data mining. In [34], the authors conducted a study on tourist market segmentation with linear and non-linear techniques. They used SOM neural networks for segmenting tourist markets and backpropagation neural networks for tourists’ classification. In [33], the authors developed a method for travellers’ segmentation and choice prediction through online reviews. They used TOPSIS and machine learning techniques and evaluated the method on the travelers’ online reviews of Wellington’s hotels

in New Zealand. In a spa hotel context, the authors in [35] focused on the effect of atmospheric cues and involvement on pleasure and relaxation. Accordingly, they proposed a new model through several constructs: Atmospheric cues, Centrality to lifestyle, Pleasure, Relaxation, Satisfaction, Self-expression and Word-of-mouth. They evaluated the model presented through the thermal spa hotels. They performed the data analysis using a structural equation model approach through Partial Least Squares (PLS). They found that atmospheric cues and involvement are important antecedents of relaxation and pleasure. In addition, they found that the feeling of relaxation is more important to lead to satisfaction than pleasure. In [36], the authors conducted a study on food and beverage service quality in spa hotels. They developed an instrument and validated through a self-administrated questionnaire which was distributed to the 331 customers at four different spa hotels in Balikesir. The results of their study showed six quality dimensions: assurance and employee knowledge, healthy and attractive food, empathy, tangibles, responsiveness of service delivery, and reliability. In a study by [37], the authors focused on sustainable hotel practices and nationality, and investigated its effect on guests' satisfaction and intention to return. They performed 329 surveys and collected data from a unique upscale tourist destination in Mexico, San Miguel de Allende. They found that indicating green practices significantly impacts guests' satisfaction levels and return intentions for Americans, Mexicans, and other nationalities. In [38], the authors conducted a study on improving consumer satisfaction in green hotels. They investigated the roles of perceived warmth, corporate social responsibility motive, and perceived competence. They found that when service delivery is successful, the satisfaction is relatively higher for green hotels versus the non-green hotels. They also found the positive impact of perceived warmth and competence to be a mediator for the relationship between service outcomes and consumer satisfaction and behavioral intentions. In [39], the authors focused on the customers' satisfaction in eco-friendly hotels. They developed a new method based on multi-criteria collaborative filtering: Higher Order Singular Value Decomposition (HOSVD), ANFIS and SOM. Also, they used the Classification and Regression Tree (CART) technique to find important features in the segments identified by SOM. To achieve the research objectives, they evaluated the method on the data collected from TripAdvisor. They found that there is a positive relationship between the customers' satisfaction and quality of services in eco-friendly hotels in different customers' segments. In [40], the authors developed a method for market segmentation and travel choice prediction in spa hotels. They developed the method using SOM, HOSVD, and CART. They used SOM and HOSVD for clustering and similarity calculation tasks, CART for travel choice prediction. The method was evaluated on the data collected from TripAdvisor. The results of their study found six spa hotel customer segments including: Wellness Seekers, Steam Room Seekers, Spa Treatment Seekers, Mineral Baths Seekers, Jacuzzi Seekers, and Cosmetic Seekers.

3. Methodology

This research relies on Multi-Criteria Decision-Making (MCDM) and learning techniques, SOM, LDA, TOPSIS and ANFIS. The schematic diagram of the proposed method is shown in Figure 1. In the first step of data analysis, we used LDA to discover the main satisfaction dimensions with their relative importance from the text-based reviews. Then, we applied SOM to find the customers' segments from the numerical ratings. To rank the hotels' attributes, we applied TOPSIS on each segment and ANFIS to predict the customers' preferences based on the satisfaction dimensions and general quality aspect of hotels. It is believed that clustering by the use of SOM technique can significantly improve the effectiveness of ANFIS in predicting customers' preferences in each segment. In the following section, we briefly introduce the components of the proposed method.

3.1. Latent Dirichlet Allocation (LDA)

LDA was proposed in [20] as an unsupervised machine learning technique for feature extraction. LDA is considered as a generative probabilistic model of a corpus. This technique is widely used in topic modelling, which is able to discover the principal topics from big data, massive volumes of

unstructured text data. In fact, LDA can be used to effectively discover a mixture of topics from the huge number of documents (i.e., online reviews). A mixture of topics can be those aspects which influence travellers’ satisfaction on the service quality of a specific hotel. In this study, we use LDA to identify an optimum number of dimensions (also referred to as “topic” in LDA literature) with the related words in each dimension from the travellers’ online reviews.

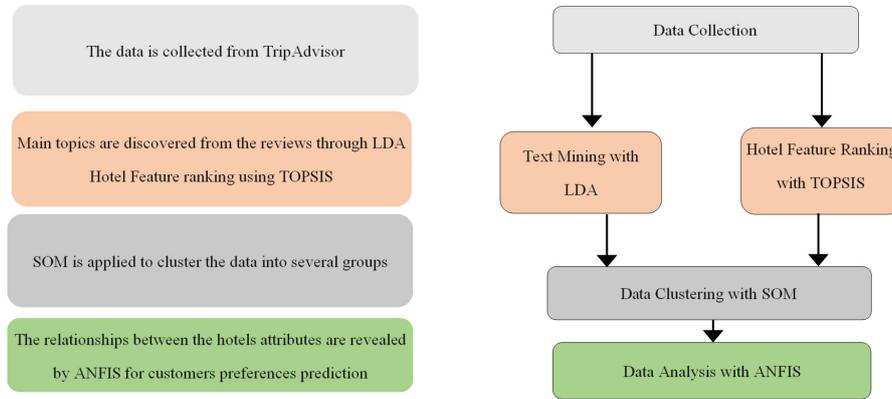


Figure 1. The schematic diagram of the proposed method.

3.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Fuzzy logic and neural network are widely used in prediction problems. ANFIS is a useful technique based on fuzzy logic and neural network approaches [27]. It takes the advantages of both fuzzy set theory in applying rule-based systems and neural networks in automatic learning from data. A fuzzy inference system in the ANFIS technique consists of if-then rules, membership functions, inference mechanism (called fuzzy reasoning), and couples of input-output. In Figure 2, a structure of a fuzzy inference system is presented. As seen from this figure, in the first step, the inputs are fuzzified to produce their degrees of truth. In the second step, the degree of truth of the consequents is obtained by combining this information through inference rules. In the last step, final output is obtained by defuzzification.

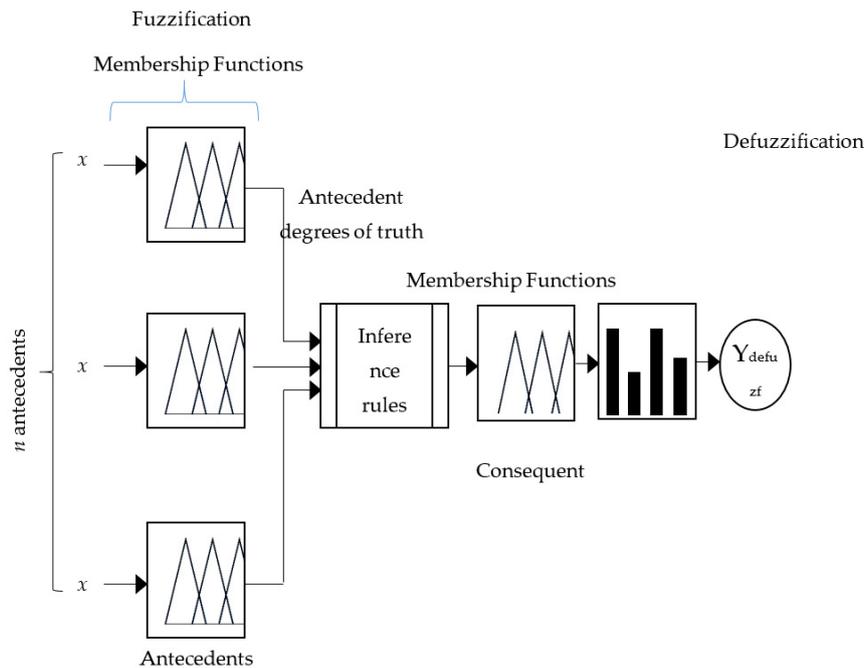


Figure 2. A structure of a fuzzy inference system for revealing customer satisfaction.

3.3. TOPSIS

TOPSIS is an MCDM technique widely used for decision-making problems [24,41]. As a compensatory aggregation method, this technique is used to select the best based on hard cut-offs. In TOPSIS, m attributes for n candidate networks can be presented in the form of a matrix $A_{(n \times m)}$ as shown in Equation (1). Therefore, the procedure for TOPSIS is presented in Figure 3.

$$A_i = \begin{pmatrix} x_{11} & x_{12} & x_{1m} \\ x_{21} & x_{22} & x_{2m} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{nm} \end{pmatrix} \quad (1)$$

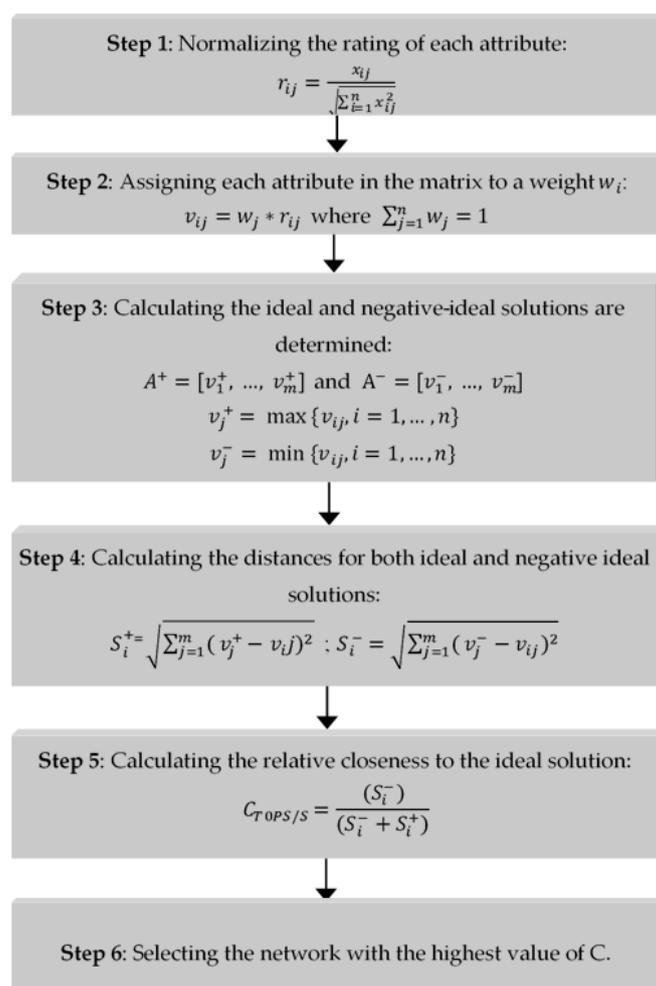


Figure 3. TOSIS procedure for criteria selection in green hotels.

4. Data Analysis

4.1. Data Collection

This research has considered TripAdvisor for data collection [23,42]. The data were collected from the hotels' webpages, which have been provided by TripAdvisor. A crawler has been designed and the data were collected through the URLs of Malaysian hotels' information. The crawler was designed to obtain the main information such as hotel information, traveller information, the trip information, and the travellers' ratings and reviews on the hotels from TripAdvisor.

By the use of the crawler, in total, we collected 17,024 records from 152 hotels (4-star and 5-star hotels). The data were preprocessed and the useless records have been removed from the database. In addition, in this stage, we removed non-English reviews from the datasets. Furthermore, the records which do not include the ratings on the hotels features (e.g., check-in/front desk, value (cost-benefit), sleep quality, service, cleanliness, rooms, location) have been removed from the database. In addition, we have eliminated those tuples which do not provide the hotels' and travel information (e.g., travelling groups). We consider two types of green hotels, green hotels with spa services and green hotels with non-spa services, for the travellers of four main groups, Travelled with friends, Solo travelers, Travelled with family and Travelled as a couple. In Table A1 of Appendix A, full information on the travellers' ratings on seven criteria of hotels is presented. In Table 1, we have provided the information of four travelling groups in 152 green hotels with spa and non-spa services.

Table 1. Full information of travelling groups and green hotels with spa and non-spa services.

Hotel Type	Traveling Type				Total
	Travelled as a couple	Travelled solo	Travelled with family	Travelled with friends	
Green Hotels with Non-Spa Services	261	204	2610	3774	6849
Green Hotels with Spa Services	4528	3236	1861	550	10,175
Total	4789	3440	4471	4324	17,024

4.2. Results and Discussions

We first applied LDA for discovering the main dimensions of satisfaction from the text-based reviews. The results are presented in Figures 4 and 5, respectively, for green hotels with non-spa and spa services. These figures reveal the relative importance of each dimension/topic (satisfaction dimension) for two types of hotels in four traveling groups, Travelled with Friends, Solo Travelers, Travelled with Family and Travelled as a Couple. The main satisfaction dimensions were: Spa treatments, Healing water, Mineral baths, Foot massage, Wellness, Sauna, Stress, Refreshing, Face treatment, Body treatment, Useful for the skin, Steam room, Ice cold water, Infrared sauna, Hot beds, Warm Jacuzzis, Warm atmosphere, Mineral water and Indoor pool. The relative importance of each dimension for green hotels with non-spa services is provided in Table A2 of Appendix A.

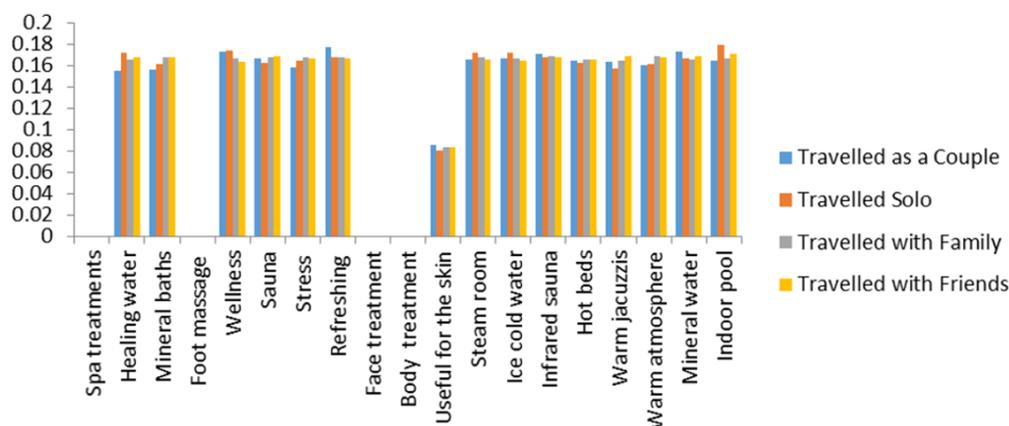


Figure 4. Travelling groups and satisfaction dimensions for green hotels with non-spa services.

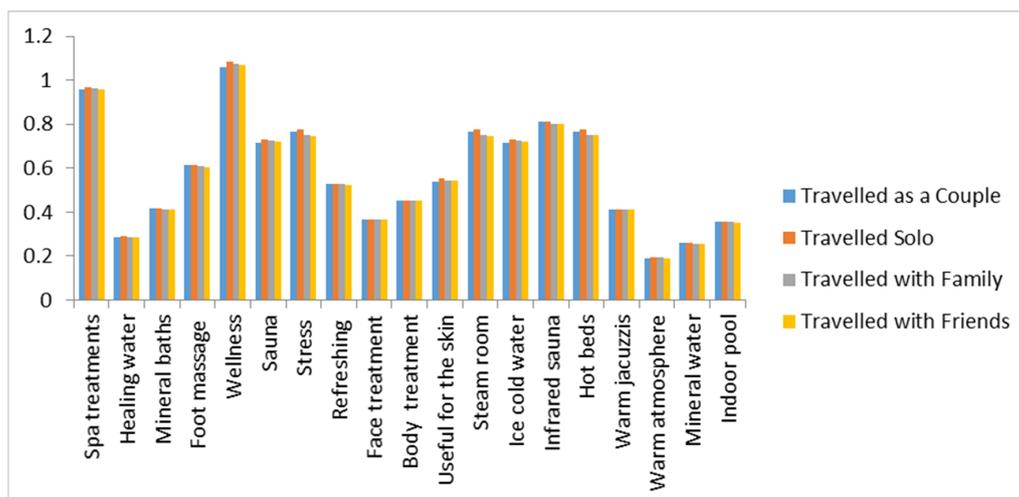


Figure 5. Travelling groups and satisfaction dimensions for green hotels with spa services.

The data were divided into two main groups, green hotels with spa and green hotels with non-spa services. Then, we applied SOM clustering for each group. The learning rate for SOM was set to 0.05. In addition, the coefficient of determination (R^2) values [43] for SOM map quality were, respectively, 0.821 and 0.845 for green hotels with spa and green hotels with non-spa, indicating that high quality of segments have been generated by SOM. For both green spa and non-spa hotels, data were clustered in four segments with least similarity between the segments. Segment 1, Segment 2, Segment 3, and Segment 4 included, respectively, 4004 (39.4%), 2586 (25.4%), 1179 (11.6%), and 2406 (23.6%) records for green spa hotels, and 1686 (24.6%), 2680 (39.1%), 698 (10.2%), and 1785 (26.1%) records for green non-spa hotels.

In Figures 6a–d and 7a–d, we provided the frequency of overall ratings on green spa and non-spa hotels based on four traveling groups in four segments. Overall, in four segments of green spa hotels, it can be found that the majority of the travellers have received high-level service quality, especially in two groups, Travelled as a couple and Travelled solo. In addition, it can be found that, in four segments of green non-spa hotels, the satisfaction level is relatively lower compared to the green spa hotels. These segments are further analyzed using Adaptive Neuro-Fuzzy Inference System, as a neuro-fuzzy system, to reveal the relative importance of each dimension on the travellers' satisfaction.

In addition to the above analysis, we applied the TOPSIS technique on the numerical ratings for each travelling group (Travelled with Friends, Solo Travelers, Travelled with Family and Travelled as a Couple) for green hotels with spa and non-spa services. The data of each cluster were considered in the matrices for TOPSIS analysis. Each value of these matrices considered a value ranged from 1 to 5. The TOPSIS technique is applied to rank the hotels' criteria selection in four segments. An example of ratings on 7 criteria of hotels is presented in Figure 8. The results of TOPSIS for ranking seven criteria of spa green hotels and non-spa green hotels in four segments are respectively presented in Tables 2 and 3.

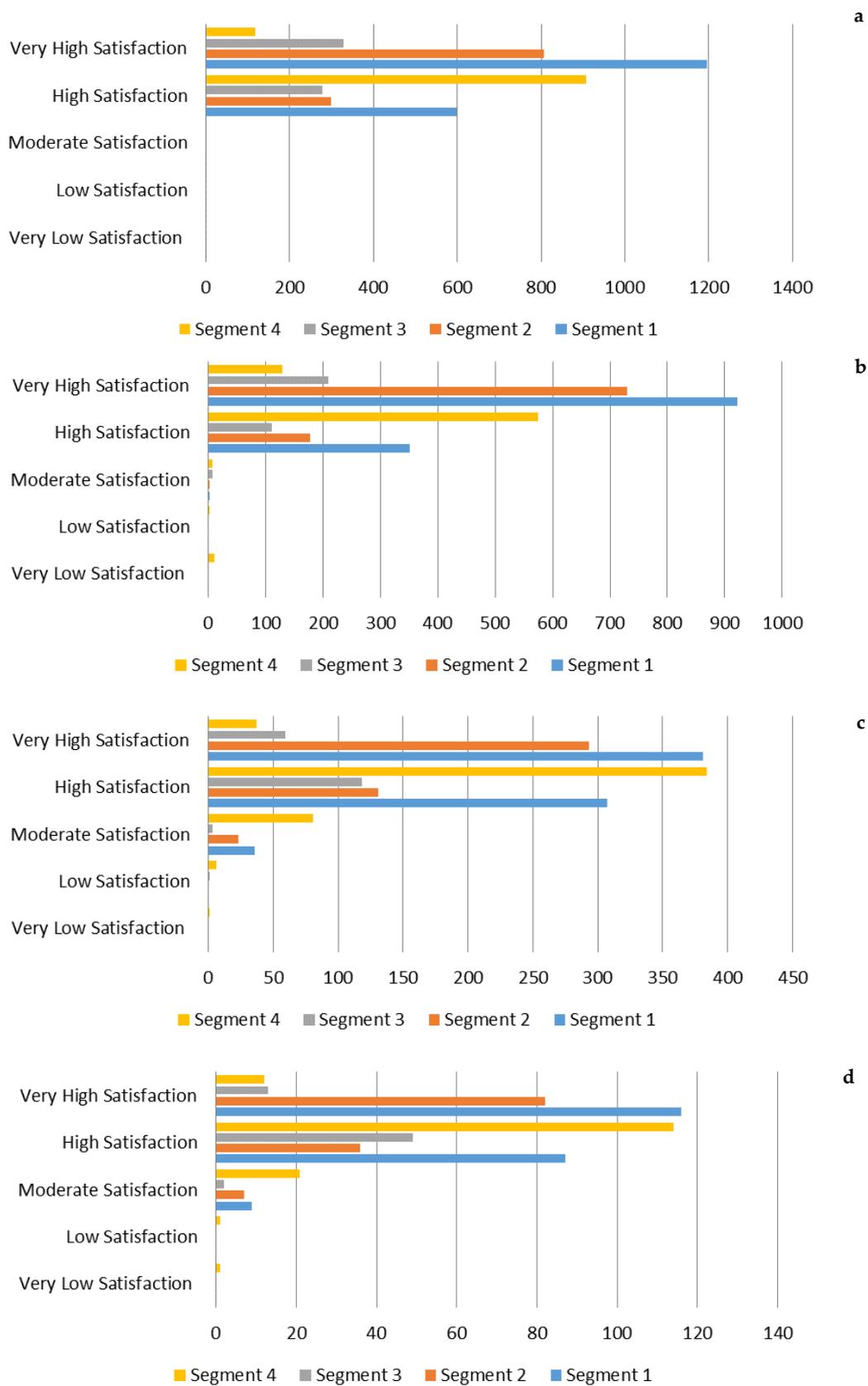


Figure 6. Travellers’ overall satisfaction in green hotels with spa services in four segments for (a) Travelled as a Couple, (b) Travelled Solo, (c) Travelled with Family, and (d) Travelled with Friends.

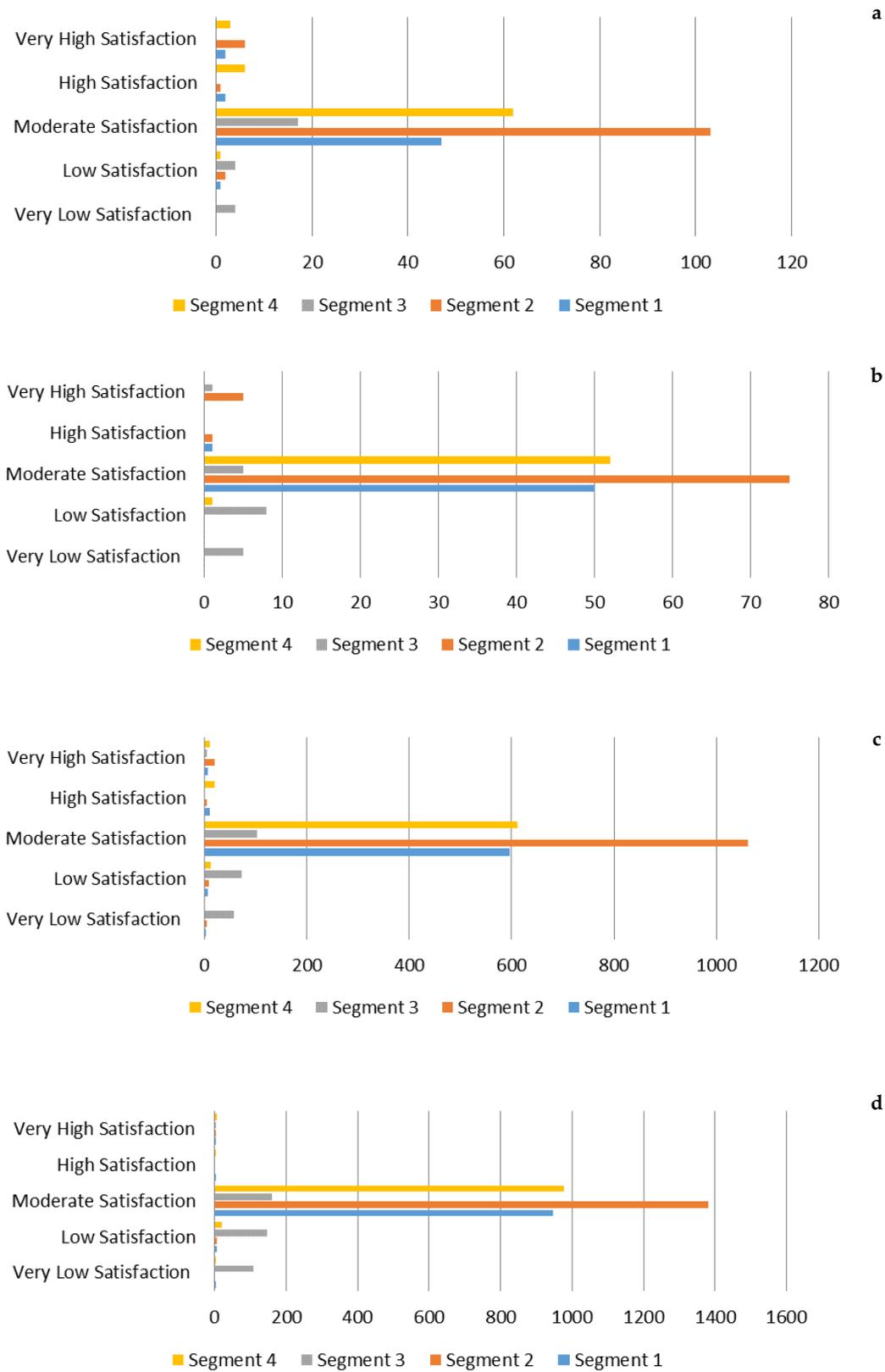


Figure 7. Travellers' overall satisfaction in green hotels with non-spa services in four segments for (a) Travelled as a Couple, (b) Travelled Solo, (c) Travelled with Family, and (d) Travelled with Friends.

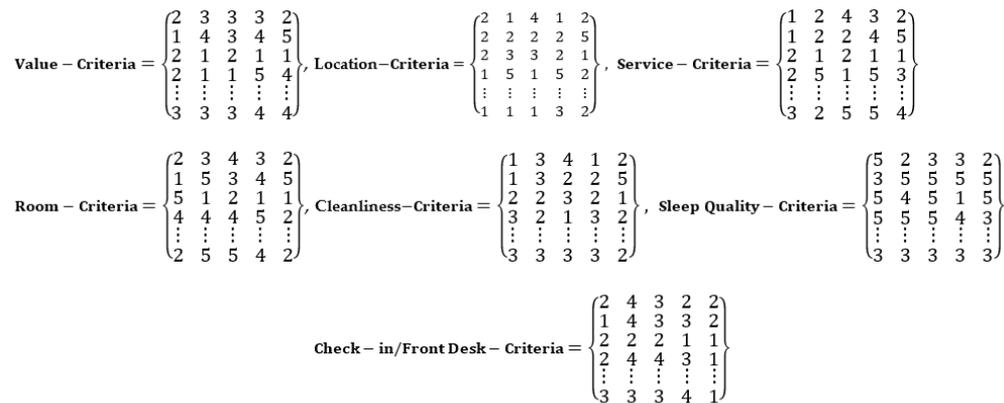


Figure 8. Example of users' ratings on the hotels' criteria.

Table 2. TOPSIS results for green hotels with spa services.

Travelled with Friends	Segment 1	Segment 2	Segment 3	Segment 4
Criteria	TOPSIS Weight			
Value	0.838	0.706	0.155	0.487
Room	0.426	0.560	0.500	0.367
Location	0.633	0.627	0.230	0.503
Cleanliness	0.791	0.467	0.257	0.337
Check-in/Front Desk	0.616	0.580	0.809	0.297
Service	0.507	0.673	0.809	0.546
Sleep Quality	0.284	0.615	0.569	0.717
Solo Travelers				
Criteria	TOPSIS Weight			
Value	0.507	0.266	0.563	0.411
Room	0.510	0.904	0.271	0.801
Location	0.606	0.438	0.586	0.595
Cleanliness	0.408	0.582	0.636	0.154
Check-in/Front Desk	0.889	0.534	0.823	0.303
Service	0.472	0.255	0.344	0.644
Sleep Quality	0.527	0.852	0.392	0.339
Travelled with Family				
Criteria	TOPSIS Weight			
Value	0.361	0.183	0.763	0.497
Room	0.459	0.658	0.668	0.468
Location	0.403	0.225	0.078	0.741
Cleanliness	0.983	0.692	0.320	0.736
Check-in/Front Desk	0.731	0.780	0.512	0.573
Service	0.518	0.667	0.821	0.592
Sleep Quality	0.965	0.536	0.333	0.820
Travelled as a Couple				
Criteria	TOPSIS Weight			
Value	0.199	0.321	0.554	0.182
Room	0.701	0.491	0.434	0.553
Location	0.667	0.557	0.307	0.488
Cleanliness	0.512	0.415	0.090	0.550
Check-in/Front Desk	0.495	0.129	0.630	0.342
Service	0.639	0.260	0.221	0.370
Sleep Quality	0.158	0.507	0.541	0.460

Table 3. TOPSIS results for green hotels with non-spa services.

Travelled with Friends				
Criteria	TOPSIS Weight			
Value	0.200	0.340	0.112	0.392
Room	0.470	0.591	0.419	0.171
Location	0.580	0.393	0.214	0.454
Cleanliness	0.524	0.286	0.057	0.199
Check-in/Front Desk	0.431	0.532	0.412	0.362
Service	0.448	0.222	0.634	0.237
Sleep Quality	0.106	0.169	0.382	0.252
Solo Travelers				
Criteria	TOPSIS Weight			
Value	0.472	0.178	0.364	0.299
Room	0.191	0.337	0.577	0.494
Location	0.455	0.276	0.159	0.096
Cleanliness	0.269	0.342	0.122	0.378
Check-in/Front Desk	0.438	0.361	0.437	0.612
Service	0.176	0.359	0.284	0.262
Sleep Quality	0.345	0.611	0.300	0.399
Travelled with Family				
Criteria	TOPSIS Weight			
Value	0.297	0.290	0.326	0.338
Room	0.348	0.198	0.245	0.136
Location	0.115	0.176	0.466	0.345
Cleanliness	0.124	0.197	0.395	0.287
Check-in/Front Desk	0.117	0.337	0.247	0.150
Service	0.391	0.463	0.232	0.572
Sleep Quality	0.606	0.264	0.393	0.363
Travelled as a Couple				
Criteria	TOPSIS Weight			
Value	0.485	0.217	0.055	0.499
Room	0.344	0.257	0.428	0.310
Location	0.228	0.274	0.431	0.302
Cleanliness	0.372	0.147	0.333	0.209
Check-in/Front Desk	0.056	0.253	0.475	0.109
Service	0.425	0.394	0.349	0.310
Sleep Quality	0.193	0.136	0.123	0.321

In the last step of data analysis, the data of each cluster were used in the Adaptive Neuro-Fuzzy Inference System to reveal the importance level of each factor on the travellers' satisfaction level. In addition, the importance of spa services on travellers' satisfaction level in four travelling groups, Travelled with Friends, Solo Travelers, Travelled with Family, and Travelled as a Couple, was assessed by the Adaptive Neuro-Fuzzy Inference System in each cluster. To do so, a Sugeno-type fuzzy inference system was developed and the data for both green spa and non-spa hotels in four segments were trained by the Adaptive Neuro-Fuzzy Inference System. In the Adaptive Neuro-Fuzzy Inference System, we used a hybrid learning algorithm, which is a combination of least-squares algorithm and gradient descent algorithm for parameters identification in learning the prediction models. This approach also converges much faster than other learning approaches. In addition, we used a 5-layer Adaptive Neuro-Fuzzy Inference System in constructing the prediction models. The network architecture and model parameters in the 5-layer Adaptive Neuro-Fuzzy Inference System were: 200 training epochs, four training and four test datasets for each type of hotels, and Gaussian membership functions.

The results of ANFIS are presented in Figures 9–12. The results of revealing the impact of service quality on travellers' overall satisfaction showed that, in general, the level of satisfaction in green

spa hotels is higher than green non-spa hotels. In all segments, the results reveal that the groups of travellers are more satisfied with the quality of services in green hotels with spa in relation to the green hotels with non-spa. The results obtained from the service quality ratings and travellers' experiences in their stay in hotels show that the hotels with spa services can contribute to high levels of satisfaction.

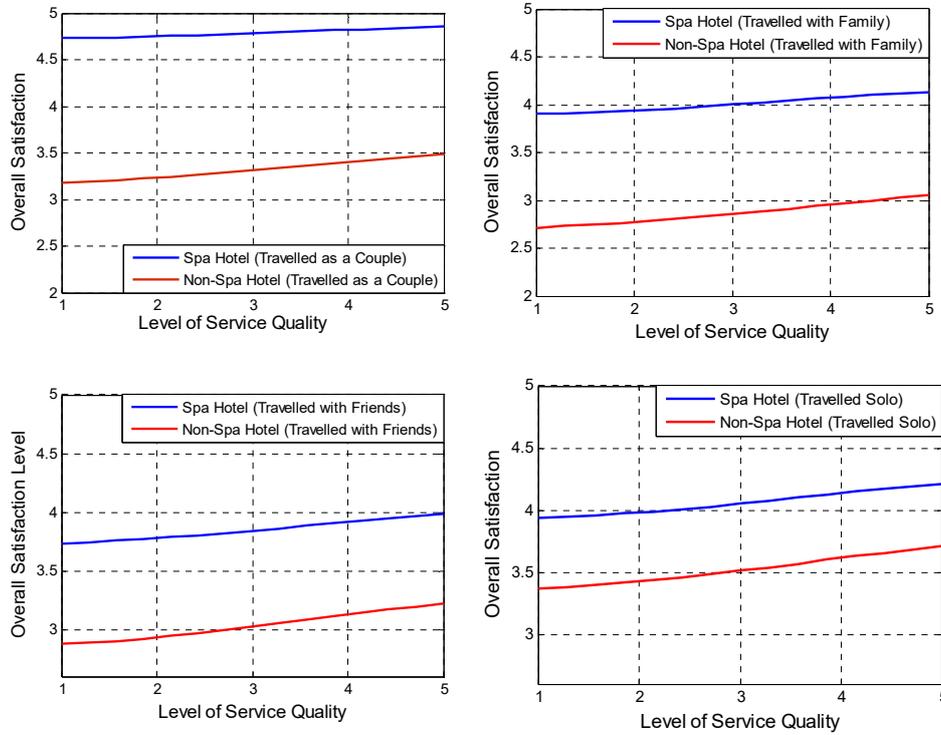


Figure 9. Overall satisfaction level in green hotels with spa and non-spa services in Segment 1.

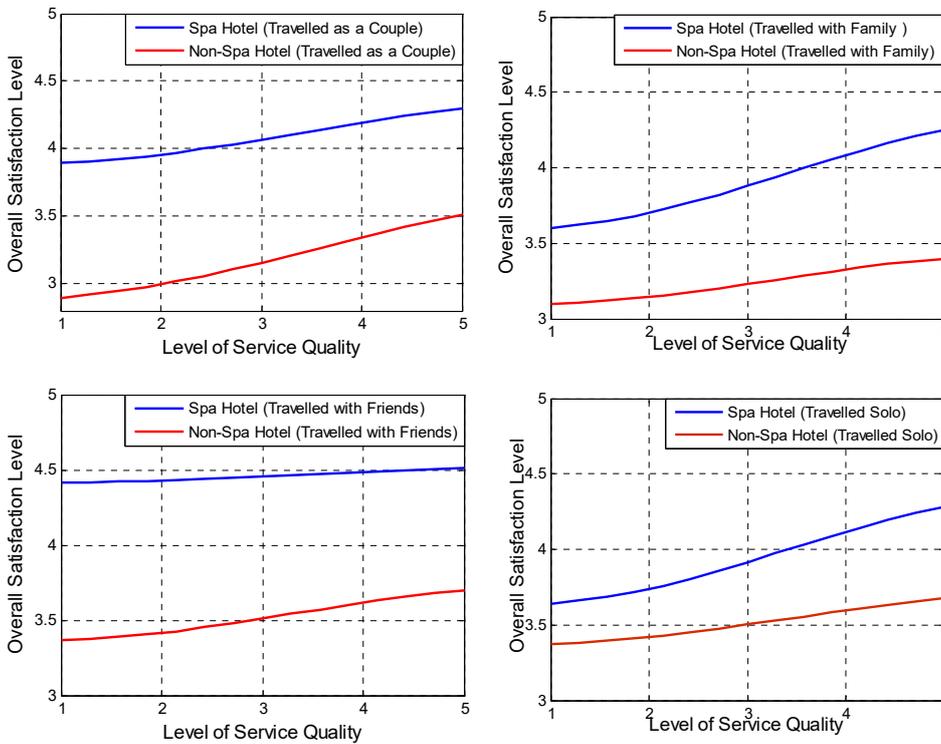


Figure 10. Overall satisfaction level in green hotels with spa and non-spa services in Segment 2.

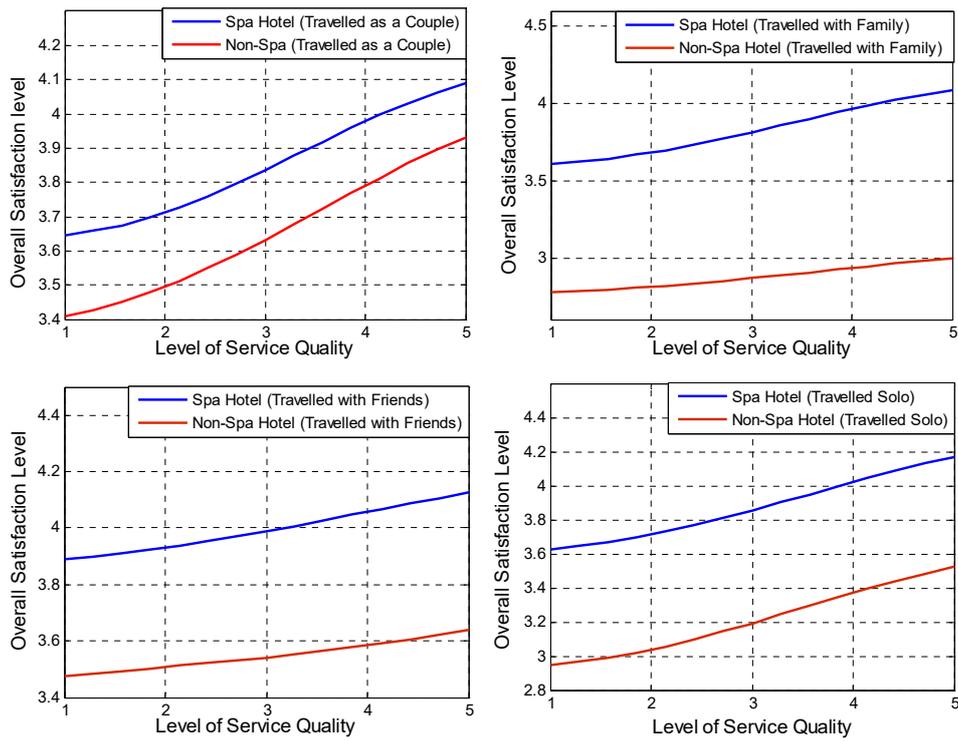


Figure 11. Overall satisfaction level in green hotels with spa and non-spa services in Segment 3.

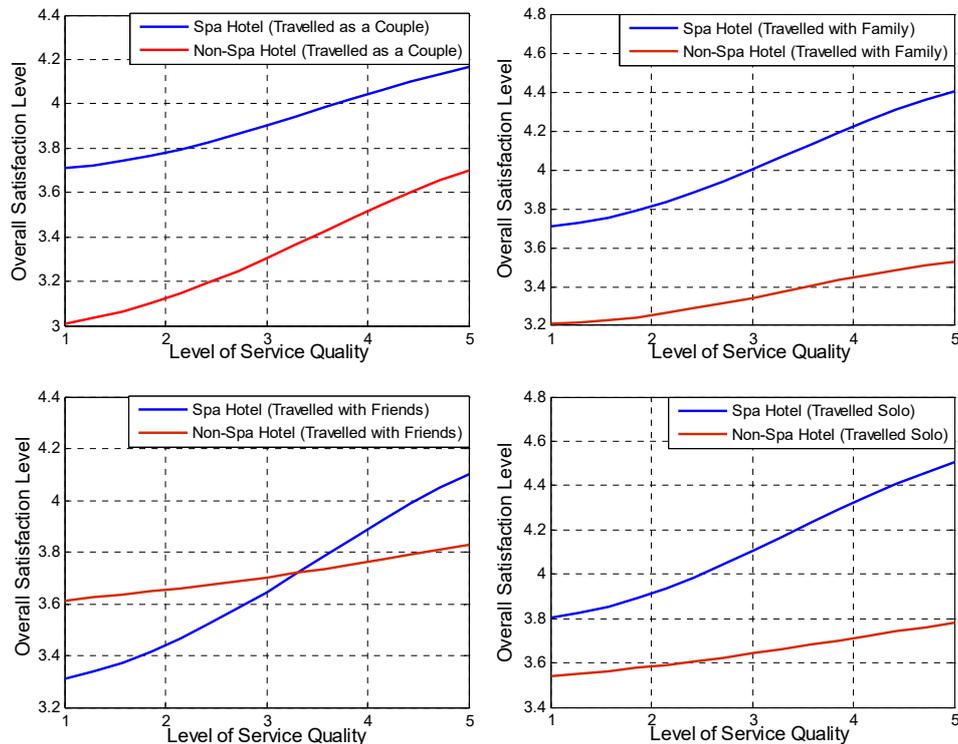


Figure 12. Overall satisfaction level in green hotels with spa and non-spa services in Segment 4.

To show the effectiveness of using clustering and the ANFIS model in the prediction of customers' preferences, we performed several experiments on the data collected from TripAdvisor and compared them with the other supervised learning methods. Specifically, we selected the Neural Network (NN), Support Vector Regression (SVR), and Multiple Linear Regression (MLR) techniques for methods comparison. The experiments were performed using a tenfold cross-validation approach. The results

for method comparisons were provided for R^2 and Mean Absolute Error (MAE) for each method. The results revealed that the method which uses clustering with the aid of the SOM and ANFIS techniques (MAE = 0.073; R^2 = 0.923) has provided better prediction accuracy in relation to the MLR (MAE = 0.273; R^2 = 0.725), SVR (MAE = 0.123; R^2 = 0.856), and NN (MAE = 0.193; R^2 = 0.807) learning techniques.

5. Discussion

Sustainability has been found to be an important issue among tourism and hospitality stakeholders and become an increasingly popular field of research since the late 1980s [44]. Sustainability is viewed as holding considerable promise to address the problems of tourism negative impacts. The studies on sustainable tourism find that consumers are nowadays more demanding than before considering hotels with green initiatives and features [45]. Previous studies also show that there is a positive relationship between the customers' satisfaction and behavioral intentions [36,38]. Furthermore, the findings of previous research on tourism reveal the positive impact of green practice on customers' satisfaction and loyalty [46].

Customer segmentation has been an important issue in marketing, especially in tourism [39]. It helps the managers effectively discover the customers' satisfaction level from their services and detect their choice preferences [40]. Previous studies on customers' segmentation and preference prediction are mainly developed through survey data [47]. However, it is found that the use of machine learning techniques can be more effective in big data analysis and customer segmentation [48]. This has been further investigated in previous studies with the aid of multi-criteria decision-making approaches [33,49]. In a tourism context, this issue was also extensively investigated [33,39,50]. However, this issue is fairly investigated in tourism and hospitality through social big data. In addition, the use of machine learning techniques to analyze social big data in eco-friendly hotels for customers' preference learning and segmentation is investigated in a few number of studies.

Accordingly, this research focused on the use of machine learning and multi-criteria decision-making techniques to develop a new method for customer segmentation in eco-friendly hotels. The method was evaluated on the travellers' online reviews on 152 hotels in Malaysia. From a marketing perspective, the results of our study demonstrated that green initiatives lead to greater consumer satisfaction, which is in line with previous findings reporting the positive impact of eco-friendly hotels on customers' satisfaction [39]. In addition, our findings demonstrated that the level of satisfaction in green spa hotels is higher than green non-spa hotels. Interestingly, the results revealed that, in all travelling groups (Travelled with Friends, Solo Travelers, Travelled with Family and Travelled as a Couple), there is a positive relationship between the quality of services in green hotels and their satisfaction level. It was further demonstrated that the satisfaction levels in these groups for spa green hotels are higher than the satisfaction levels in the non-spa green hotels. From a methodological perspective, the results showed that the use of machine learning techniques with the aid of multi-criteria decision-making approaches can better predict the customers' preferences in eco-friendly hotels in relation to the method which solely relies on predictive learning techniques. In addition, the results revealed that the methods which use a clustering approach can better reveal the customers' preferences in eco-friendly hotels. This shows the effectiveness of clustering in customer segmentation for customers' preferences prediction [39]. This was confirmed with the comparisons of the proposed method with the other predictive learning techniques in the literature.

6. Conclusions

This paper investigated customer satisfaction in green hotels through travellers' online review analysis. We used machine learning techniques to analyze the online text-based reviews and numerical data analysis. SOM was used for travelers' segmentation in both green spa and non-spa hotels. LDA was used for the identification of the main satisfaction dimensions (topics) from the text-based reviews. We also used TOPSIS to rank the main criteria for green hotel selection from numerical reviews. The neuro-fuzzy system was then used to reveal the relative importance of the hotel features on the

customers' satisfaction and find the importance role of spa services and its impact on the customers' satisfaction. We investigated the role of spa services on the travellers' overall satisfaction in four main travelling groups, Travelled with Friends, Solo Travelers, Travelled with Family and Travelled as a Couple. The data were collected from Malaysian green spa and non-spa hotels in the TripAdvisor platform. The results showed that, in general, the travellers' satisfaction was relatively higher than in green spa hotels in relation to the green non-spa hotels. This result was proved in four travelling groups, Travelled with Friends, Solo Travelers, Travelled with Family and Travelled as a Couple. In our experimental analysis, we observed that travellers are relatively more satisfied in green spa hotels and accordingly provide a higher degree of overall ratings to the hotels with spa services.

This research has some limitations. First, this study has considered only the data from TripAdvisor to investigate the role of spa services on customers' satisfaction. The data from other social networking sites such as trivago.com, booking.com, and trip.com can further confirm the results of this study. Second, this study only considered the Malaysian green spa and non-spa hotels to validate the proposed hypothesis. The data from hotels registered in TripAdvisor for other developed and developing countries may better reveal the significance of spa services on customers' satisfaction. Third, although our method was implemented to assess the travellers' satisfaction in green hotels, the proposed method can also be implemented for other types of hotels (e.g., non-green hotels with spa services) and then comparisons are made for the travellers' satisfaction levels in different hotels (i.e., comparisons for travellers' satisfaction levels in green hotels with spa services and non-green hotels with spa services). Fourth, four travelling groups have been considered for hypothesis evaluation. Future studies can investigate the role of spa services on other travelling groups such as business travellers. Fifth, this study does not investigate the role of spa services on the travelers' satisfaction from the gender perspective. This can be an interesting investigation for future studies in the tourism context. Sixth, our method was based on non-incremental learning techniques. The use of deep learning approaches, Random Forest (RF) and incremental learning or the development of our method with incremental updates, especially in LDA and neuro-fuzzy learning techniques, can be effective in online reviews analysis. This limitation could be a rich avenue for future research. Besides the TOPSIS method, we can also consider group decision-making (GDM) and other multiple criteria decision-making methods such as the double normalization-based multiple aggregation (DNMA) method [51] and the gained and lost dominance score (GLDS) method [52], to solve such problems. Finally, multicollinearity phenomenon may occur due to the high correlations among the features of dataset which should be solved through appropriate learning techniques.

Author Contributions: The research was designed and performed by M.N. and A.M. The data was collected and analyzed by M.N. and A.M. The paper was written by M.N., A.M. and H.L., and finally checked and revised by H.A., A.A.M. and W.A. All authors read and approved the final manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 71771156 and 71971145, and the 2018 Key Project of the Key Research Institute of Humanities and Social Sciences in Sichuan Province, grant number Xq18A01 and LYC18-02.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Full information of travellers' ratings on seven criteria of green hotels with spa services.

		Overall Rating					Total
		1	2	3	4	5	
Value	1	1	0	26	95	196	318
	2	0	2	37	136	217	392
	3	44	58	210	446	680	1438
	4	108	210	923	4170	5047	10,458
	5	39	41	394	1176	2768	4418
Total		192	311	1590	6023	8908	17,024
		Overall Rating					Total
		1	2	3	4	5	
Room	2	8	7	10	2	21	48
	3	19	75	373	954	1416	2837
	4	73	97	798	3529	3993	8490
	5	92	132	409	1538	3478	5649
Total		192	311	1590	6023	8908	17,024
		Overall Rating					Total
		1	2	3	4	5	
Location	1	10	8	9	30	52	109
	2	15	12	38	55	111	231
	3	23	30	155	716	738	1662
	4	123	220	1031	3446	5224	10,044
	5	21	41	357	1776	2783	4978
Total		192	311	1590	6023	8908	17,024
		Overall Rating					Total
		1	2	3	4	5	
Cleanliness	1	0	1	17	71	138	227
	2	1	1	36	117	196	351
	3	1	6	99	257	405	768
	4	159	250	902	3324	3518	8153
	5	31	53	536	2254	4651	7525
Total		192	311	1590	6023	8908	17,024
		Overall Rating					Total
		1	2	3	4	5	
Check-in/Front Desk	1	35	22	16	81	132	286
	2	23	29	47	72	171	342
	3	26	38	175	285	451	975
	4	104	204	1138	4427	5665	11,538
	5	4	18	214	1158	2489	3883
Total		192	311	1590	6023	8908	17,024

Table A1. Cont.

	Overall Rating					Total	
	1	2	3	4	5		
Service	1	73	43	17	21	45	199
	2	31	87	80	145	182	525
	3	17	71	330	656	879	1953
	4	67	94	762	3454	3938	8315
	5	4	16	401	1747	3864	6032
Total	192	311	1590	6023	8908	17,024	
Sleep Quality	Overall Rating					Total	
	1	2	3	4	5		
	1	156	15	53	27	64	315
	2	12	221	83	56	80	452
	3	12	30	95	74	411	622
	4	5	34	530	2416	1861	4846
5	7	11	829	3450	6492	10,789	
Total	192	311	1590	6023	8908	17,024	

Table A2. Full information of travelling groups and satisfaction dimensions for green hotels with non-spa services.

Satisfaction Dimensions	Traveling Type			
	Travelled as a Couple	Travelled Solo	Travelled with Family	Travelled with Friends
	Mean	Mean	Mean	Mean
Spa treatments	0.0000	0.0000	0.0000	0.0000
Healing water	0.1544	0.1712	0.1659	0.1678
Mineral baths	0.1558	0.1612	0.1672	0.1670
Foot massage	0.0000	0.0000	0.0000	0.0000
Wellness	0.1724	0.1734	0.1663	0.1634
Sauna	0.1669	0.1625	0.1672	0.1681
Stress	0.1581	0.1647	0.1676	0.1667
Refreshing	0.1768	0.1673	0.1675	0.1663
Face treatment	0.0000	0.0000	0.0000	0.0000
Body treatment	0.0000	0.0000	0.0000	0.0000
Useful for the skin	0.0850	0.0804	0.0828	0.0828
Steam room	0.1651	0.1721	0.1670	0.1657
Ice cold water	0.1663	0.1716	0.1666	0.1642
Infrared sauna	0.1707	0.1675	0.1686	0.1678
Hot beds	0.1645	0.1624	0.1655	0.1650
Warm jacuzzis	0.1636	0.1566	0.1648	0.1684
Warm atmosphere	0.1604	0.1617	0.1686	0.1670
Mineral water	0.1732	0.1664	0.1658	0.1687
Indoor pool	0.1641	0.1787	0.1661	0.1710

References

1. Rahman, I.; Reynolds, D. The influence of values and attitudes on green consumer behavior: A conceptual model of green hotel patronage. *Int. J. Hosp. Tour. Adm.* **2019**, *20*, 47–74. [[CrossRef](#)]
2. Gupta, A.; Dash, S.; Mishra, A. All that glitters is not green: Creating trustworthy ecofriendly services at green hotels. *Tour. Manag.* **2019**, *70*, 155–169. [[CrossRef](#)]
3. Trang, H.L.T.; Lee, J.-S.; Han, H. How do green attributes elicit pro-environmental behaviors in guests? The case of green hotels in Vietnam. *J. Travel Tour. Mark.* **2019**, *36*, 14–28. [[CrossRef](#)]
4. Jiang, Y.; Gao, Y. (Lisa) Factors that Influence Potential Green Hotel Customers' Decision-making Process—Evidence from China. *J. China Tour. Res.* **2019**, *15*, 455–477. [[CrossRef](#)]
5. Yadav, R.; Balaji, M.; Jebarajakirthy, C. How psychological and contextual factors contribute to travelers' propensity to choose green hotels? *Int. J. Hosp. Manag.* **2019**, *77*, 385–395. [[CrossRef](#)]
6. Gil-Soto, E.; Armas-Cruz, Y.; Morini-Marrero, S.; Ramos-Henriquez, J.M. Hotel guests' perceptions of environmental friendly practices in social media. *Int. J. Hosp. Manag.* **2019**, *78*, 59–67. [[CrossRef](#)]
7. Cheng, J.C.-H.; Chiang, A.-H.; Yuan, Y.; Huang, M.-Y. Exploring Antecedents of Green Tourism Behaviors: A Case Study in Suburban Areas of Taipei, Taiwan. *Sustainability* **2018**, *10*, 1928. [[CrossRef](#)]
8. Yfantidou, G.; Matarazzo, M. The future of sustainable tourism in developing countries. *Sustain. Dev.* **2017**, *25*, 459–466. [[CrossRef](#)]
9. Han, H.; Ryu, K. Moderating Role of Personal Characteristics in Forming Restaurant Customers' Behavioral Intentions: An Upscale Restaurant Setting. *J. Hosp. Leis. Mark.* **2007**, *15*, 25–54. [[CrossRef](#)]
10. Jani, D.; Han, H. Personality, social comparison, consumption emotions, satisfaction, and behavioral intentions: How do these and other factors relate in a hotel setting? *Int. J. Contemp. Hosp. Manag.* **2013**, *25*, 970–993. [[CrossRef](#)]
11. Choi, Y.; Kim, J.; Lee, C.K.; Hickerson, B. The role of functional and wellness values in visitors' evaluation of spa experiences. *Asia Pac. J. Tour. Res.* **2015**, *20*, 263–279. [[CrossRef](#)]
12. Sharma, P.; Nayak, J.K. Testing the role of tourists' emotional experiences in predicting destination image, satisfaction, and behavioral intentions: A case of wellness tourism. *Tour. Manag. Perspect.* **2018**, *28*, 41–52. [[CrossRef](#)]
13. Han, H.; Kiatkawsin, K.; Jung, H.; Kim, W. The role of wellness spa tourism performance in building destination loyalty: The case of Thailand. *J. Travel Tour. Mark.* **2017**, *35*, 595–610. [[CrossRef](#)]
14. Oliver, R.L. Emotional expression in the satisfaction response. In *Satisfaction: A Behavioral Perspective on the Consumer*; Routledge: New York, NY, USA, 1997; pp. 291–325.
15. Han, H.; Kiatkawsin, K.; Kim, W.; Lee, S. Investigating customer loyalty formation for wellness spa: Individualism vs. collectivism. *Int. J. Hosp. Manag.* **2017**, *67*, 11–23. [[CrossRef](#)]
16. González, M.E.A.; Comesaña, L.R.; Brea, J.A.F. Assessing tourist behavioral intentions through perceived service quality and customer satisfaction. *J. Bus. Res.* **2007**, *60*, 153–160. [[CrossRef](#)]
17. Guo, Y.; Barnes, S.J.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [[CrossRef](#)]
18. Lu, W.; Stepchenkova, S. Ecotourism experiences reported online: Classification of satisfaction attributes. *Tour. Manag.* **2012**, *33*, 702–712. [[CrossRef](#)]
19. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [[CrossRef](#)]
20. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
21. Yu, J.; Mohan, S.; Wong, W.K.; Putthividhya, D. Latent dirichlet allocation based diversified retrieval for e-commerce search. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, NY, USA, 24–28 February 2014.
22. Liu, Y.; Teichert, T.; Rossi, M.; Li, H.; Hu, F. Big data for big insights: Investigating language-specific drivers of hotel satisfaction with user-generated reviews. *Tour. Manag.* **2017**, *59*, 554–563. [[CrossRef](#)]
23. Nilashi, M.; Ibrahim, O.; Yadegaridehkordi, E.; Samad, S.; Akbari, E.; Alizadeh, A. Travelers decision making using online review in social network sites: A case on TripAdvisor. *J. Comput. Sci.* **2018**, *28*, 168–179. [[CrossRef](#)]
24. Behzadian, M.; Otaghsara, S.K.; Yazdani, M.; Ignatius, J. A state-of-the-art survey of TOPSIS applications. *Expert Syst. Appl.* **2012**, *39*, 13051–13069. [[CrossRef](#)]

25. Nilashi, M.; Ibrahim, O.B. A model for detecting customer level intentions to purchase in B2C websites using TOPSIS and fuzzy logic rule-based system. *Arab. J. Sci. Eng.* **2014**, *39*, 1907–1922. [[CrossRef](#)]
26. Fu, Z.G.; Liao, H.C. Unbalanced double hierarchy linguistic term set: The TOPSIS method for multi-expert qualitative decision making involving green mine selection. *Inf. Fusion* **2019**, *51*, 271–286. [[CrossRef](#)]
27. Jang, J.-S. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [[CrossRef](#)]
28. Jonker, J.-J.; Piersma, N.; Poel, D.V.D. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. *Expert Syst. Appl.* **2004**, *27*, 159–168. [[CrossRef](#)]
29. Chen, K.-H.; Liu, H.-H.; Chang, F.-H. Essential customer service factors and the segmentation of older visitors within wellness tourism based on hot springs hotels. *Int. J. Hosp. Manag.* **2013**, *35*, 122–132. [[CrossRef](#)]
30. Raju, C.V.L.; Narahari, Y.; Ravikumar, K. Learning dynamic prices in electronic retail markets with customer segmentation. *Ann. Oper. Res.* **2006**, *143*, 59–75. [[CrossRef](#)]
31. Florez-Lopez, R.; Ramon-Jeronimo, J.M. Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting. *Soc. Sci. Comput. Rev.* **2009**, *27*, 96–117. [[CrossRef](#)]
32. Liao, S.-H.; Chen, Y.-J.; Deng, M.-Y. Mining customer knowledge for tourism new product development and customer relationship management. *Expert Syst. Appl.* **2010**, *37*, 4212–4223. [[CrossRef](#)]
33. Ahani, A.; Nilashi, M.; Ibrahim, O. Travellers segmentation and choice prediction through online reviews: The case of Wellington’s hotels in New Zealand. *J. Soft Comput. Decis. Support Syst.* **2019**, *6*, 23–30.
34. Bloom, J.Z. Tourist market segmentation with linear and non-linear techniques. *Tour. Manag.* **2004**, *25*, 723–733. [[CrossRef](#)]
35. Loureiro, S.M.C.; Almeida, M.; Rita, P. The effect of atmospheric cues and involvement on pleasure and relaxation: The spa hotel context. *Int. J. Hosp. Manag.* **2013**, *35*, 35–43. [[CrossRef](#)]
36. Girtlioglu, I.; Jones, E.; Avcikurt, C. Measuring food and beverage service quality in spa hotels: A case study in Balikesir, Turkey. *Int. J. Contemp. Hosp. Manag.* **2014**, *26*, 183–204. [[CrossRef](#)]
37. Berezan, O.; Raab, C.; Yoo, M.; Love, C. Sustainable hotel practices and nationality: The impact on guest satisfaction and guest intention to return. *Int. J. Hosp. Manag.* **2013**, *34*, 227–233. [[CrossRef](#)]
38. Gao, Y.L.; Mattila, A.S. Improving consumer satisfaction in green hotels: The roles of perceived warmth, perceived competence, and CSR motive. *Int. J. Hosp. Manag.* **2014**, *42*, 20–31. [[CrossRef](#)]
39. Nilashi, M.; Ahani, A.; Esfahani, M.D.; Yadegaridehkordi, E.; Samad, S.; Ibrahim, O.; Sharef, N.M.; Akbari, E.; Ahnai, A. Preference learning for eco-friendly hotels recommendation: A multi-criteria collaborative filtering approach. *J. Clean. Prod.* **2019**, *215*, 767–783. [[CrossRef](#)]
40. Ahani, A.; Nilashi, M.; Ibrahim, O.; Sanzogni, L.; Weaven, S. Market segmentation and travel choice prediction in spa hotels through tripadvisor’s online reviews. *Int. J. Hosp. Manag.* **2019**, *80*, 52–77. [[CrossRef](#)]
41. Lai, Y.-J.; Liu, T.-Y.; Hwang, C.-L. Topsis for MODM. *Eur. J. Oper. Res.* **1994**, *76*, 486–500. [[CrossRef](#)]
42. Zhang, H.-Y.; Ji, P.; Wang, J.-Q.; Chen, X.-H. A novel decision support model for satisfactory restaurants utilizing social information: A case study of TripAdvisor.com. *Tour. Manag.* **2017**, *59*, 281–297. [[CrossRef](#)]
43. Sabzi, S.; Javadikia, P.; Rabani, H.; Adelkhani, A. Mass modeling of Bam orange with ANFIS and SPSS methods for using in machine vision. *Measurement* **2013**, *46*, 3333–3341. [[CrossRef](#)]
44. Higgins-Desbiolles, F. Sustainable tourism: Sustaining tourism or something more? *Tour. Manag. Perspect.* **2018**, *25*, 157–160. [[CrossRef](#)]
45. Sony, M. Managerial Perspectives on Willingness to Pay for Green Marketing: An Interpretative Phenomenological Analysis. In *Managing Diversity, Innovation, and Infrastructure in Digital Business*; IGI Global: Hershey, PA, USA, 2019; pp. 116–138.
46. Yusof, Y.; Awang, Z.; Jusoff, K.; Ibrahim, Y. The influence of green practices by non-green hotels on customer satisfaction and loyalty in hotel and tourism industry. *Int. J. Green Econ.* **2017**, *11*, 1–14. [[CrossRef](#)]
47. Guillet, B.D.; Kucukusta, D. Spa market segmentation according to customer preference. *Int. J. Contemp. Hosp. Manag.* **2016**, *28*, 418–434. [[CrossRef](#)]
48. Wu, R.-S.; Chou, P.-H. Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electron. Commer. Res. Appl.* **2011**, *10*, 331–341. [[CrossRef](#)]
49. Güçdemir, H.; Selim, H. Integrating multi-criteria decision making and clustering for business customer segmentation. *Ind. Manag. Data Syst.* **2015**, *115*, 1022–1040. [[CrossRef](#)]

50. Schuckert, M.; Liu, X.; Law, R.; Law, C.H.R. A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *Int. J. Hosp. Manag.* **2015**, *48*, 143–149. [[CrossRef](#)]
51. Liao, H.; Wu, X. DNMA: A double normalization-based multiple aggregation method for multi-expert multi-criteria decision making. *Omega* **2019**. [[CrossRef](#)]
52. Wu, X.; Liao, H. A consensus-based probabilistic linguistic gained and lost dominance score method. *Eur. J. Oper. Res.* **2019**, *272*, 1017–1027. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).