MDPI

Article

Research on the Role of Influencing Factors on Hotel Customer Satisfaction Based on BP Neural Network and Text Mining

Jiaying Wang [1,2,3]**, Zhijie Zhao** [1,2,*]**, Yang Liu** [1,2] **and Yiqi Guo** [1,2]

1. School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China; wangjiaying@uzz.edu.cn (J.W.); liuy31820@gmail.com (Y.L.); bikariri4774@gmail.com (Y.G.)
2. Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin 150028, China
3. School of Opto-Electronic Engineering, Zaozhuang University, Zaozhuang 277160, China
* Correspondence: Zhaozj@hrbcu.edu.cn

Abstract: With the flourishing development of the hotel industry, the study of customer satisfaction based on online reviews and data has become a new model. In this paper, customer reviews and ratings on Ctrip.com are used, and TF-IDF and K-means algorithms are used to extract and cluster the keywords of reviews texts. Finally, 10 first-level influencing factors of hotel customer satisfaction are determined: epidemic prevention, consumption emotion, convenience, environment, facilities, catering, target group, perceived value, price, and service. Based on backpropagation neural network and weight matrix operation, an influencing factor analysis model of hotel customer satisfaction is constructed to explore the role of these factors. The results show that consumption emotion, perceived value, epidemic prevention, target group, and convenience would significantly affect customer satisfaction, among which epidemic prevention becomes a new factor affecting customer satisfaction. Environment, facilities, catering, and service have relatively little effect on customer satisfaction, while price has the least effect. This study provides a path and method for online reviews of hotel management to improve customer satisfaction and provides a theoretical basis for the study of online reviews of hotels.

Keywords: BP neural network; text mining; online review; customer satisfaction

1. Introduction

At present, the hotel industry is very competitive, and if we want to stand out among them, customer satisfaction needs to be valued. Customer satisfaction is an important barometer for assessing hotel performance, so a deep understanding of the factors of customer satisfaction and dissatisfaction is essential for hotel management [1]. According to the service profit chain theory, customer consumption experience with a product directly determines customer satisfaction [2], which also affects the final value that customers bring to hotels. Hotel customer satisfaction refers to the contrasting results of the quality of products or services provided by hotels and customer benefit needs and has a huge impact on the image promotion of hotel brands as well as the sustainable development of competitiveness [3].

There are many studies on the customer satisfaction characteristics of hotels, and many scholars have conducted studies on hotel customer satisfaction. For example, Duan et al. studied the impact of dimensions such as service, catering, facilities, hygiene, location, and price of hotels on customer satisfaction [4]. Xiong and Xu found the influence of some factors, such as the room, network, catering, and site selection of hotels, on the comprehensive satisfaction of hotels [5]. Tsaur and Lin indicated that value, room, and service are the most concerned factors for hotels by customers [6], while Pine and Phillips pointed out that consumers have different preferences for hotels of different grades: consumers of luxury hotels have strict web requirements, while guests of economy hotels have relatively

high requirements for basic cleaning services such as toothpaste and toothbrush [7]. The above studies adopted the questionnaire method with more questions and options, which required a high theoretical basis for researchers. With the rise of review sites and travel sites, customer's decision to choose hotels is increasingly influenced by online reviews [8]. Online review text information provides many real and effective data for the study of customer satisfaction, and traditional questionnaires are no longer an obligatory way to obtain data. Scholars excavate factors affecting customer satisfaction from online review texts to find customers' real concerns [9]. Based on the online review texts to study the degree of influence of hotel characteristics on consumer satisfaction, existing studies mainly use the following methods: domain expert opinion, grammar research method, and model analysis method. The most intuitive method is to identify and evaluate the hotel characteristics based on the cognitive opinions of experts in related fields, but the expert opinions cannot represent the real experience of a large number of customers. In addition, expert opinions with strong subjectivity can suffer from many biases [10]. Grammar research methods refer to characterized words that are modified by an increasing number of adjectives, which can be speculated to be more important. With syntactic dependencies, the number of adjectives that modify the feature word is determined, and then the importance of the features is clustered for adjective recognition [11,12]. Using multivariate regression, Ganu et al. calculated the significance of each feature with feature–perspective pairs as independent variables and review ratings as dependent variables [13]. Archak et al. employed a econometric model to study the effects of characteristics' emotions on consumer willingness [14].

Some scholars have conducted research on customer satisfaction of hotels [15]; however, there are certain limitations to the existing research. First, it is not possible to determine the effect size of the influencing factors or whether the influence of the factors on customers is positive or negative. Second, most existing studies are based on text reviews and ignore customers' review ratings. Online reviews of hotels published by customers include textual reviews and review ratings, which are customer ratings of satisfaction with hotels. Geetha et al. indicated that textual reviews and review ratings are emotionally congruent [16]. Ban et al. conducted the distribution of hotel experience evaluation based on the review rating, identified the overall satisfaction level of hotel experience, and defined the review rating as hotel customer satisfaction [17]. Therefore, this paper uses review rating to represent hotel customer satisfaction.

It is very important to find out the influencing factors of hotel customer satisfaction through the text data of hotel online reviews. To analyze the impact of these factors on customer satisfaction, a rigorous machine-learning method is necessary. Artificial neural networks are represented as interconnected systems of neurons that can compute various values from the input information and can learn the intrinsic nature of patterns or processes from a dataset [18]. Recent studies have shown that neural networks are an effective alternative to traditional statistical techniques [19,20]. In particular, neural networks can provide better predictions than traditional regression methods and can be appropriately used to test large-scale data with a relatively large number of input variables [21]. This paper employs a backpropagation (BP) neural network, which is a multi-layer feed-forward neural network trained following an error backpropagation algorithm.

Therefore, starting from the online review data and review ratings of hotels, this paper uses the word separation and clustering techniques of text mining to process the hotel review text information and obtains the secondary influencing factors that affect the customer satisfaction of hotels, which contains new affecting factors. Then, according to the K-means clustering results and the Chinese customer satisfaction index (CCSI), we obtain the first-order influencing factors on hotel customer satisfaction. BP neural network and right matrix are used to construct a model of customer satisfaction influencing factors in hotels, and the effect size and influencing polarity of first-order influencing factors on customer satisfaction are analyzed. Finally, based on the research conclusions, we make targeted management recommendations for improving customer satisfaction of hotels, optimizing the service level, and enhancing the market competitiveness of hotels.

The remainder of the paper is organized in the following way. Section 2 reviews the relevant literature. Section 3 constructs an analysis model of customer satisfaction factors of hotels based on the calculation of natural language processing, BP neural network, and weight matrix and experimental analysis is conducted. Section 4 provides conclusions and theoretical and managerial implications. Finally, limitations and future research directions are given in Section 5.

2. Literature Review

2.1. Customer Satisfaction

Expectation theory states that customer satisfaction is a type of feelings that customers have after their products exceed their inner expectations, and they are disappointed and dissatisfied if evaluations of a product do not exceed their original psychological expectations [22]. Many scholars have conducted research on customer satisfaction evaluation. Among the customer satisfaction evaluation models, the most classic are the American Customer Satisfaction Index (ACSI) and the European Customer Satisfaction Index (ECSI), and most of the various satisfaction evaluation models proposed by scholars after are based on these two models for improvement and development. Tontini added an adjustment factor to the coefficient of customer satisfaction in the model, assigned greater weight to the factors that affect customer satisfaction, and realized the evaluation of customer satisfaction based on the weight of the influencing factors [23]. Shi refined the evaluation indexes in the model and reconstituted the satisfaction evaluation model system from different aspects [24]. Li and Wei proposed information satisfaction and information system satisfaction as two key factors in customer satisfaction evaluation models based on e-commerce characteristics [25]. Liu designed a customer satisfaction evaluation model of logistics enterprises using fuzzy algorithm [26]. China has also accumulated some experience with satisfaction evaluation research, with the formal implementation of CCSI in 2015. Yang and Wang improved the existing theoretical models of customer satisfaction evaluation, constructed a system of customer satisfaction evaluation using structural equations, and finally validated the model using fuzzy hierarchical analysis [27]. On the customer satisfaction evaluation method, Schneider et al. summarized the satisfaction level of customers based on a customer satisfaction index model to evaluate the customer satisfaction from a qualitative point of view [28], but their research lacks the basis of quantitative analysis. To break through the limitations of the original model and realize the quantitative research of customer satisfaction, Chen et al. used fuzzy hierarchical analysis to evaluate the customer satisfaction, determined the weight of indicators by establishing a comparison matrix, and calculated the customer satisfaction by affiliation [29]. Sun et al. proposed using the Kano model to evaluate customer satisfaction, statistically analyze the relationship between demand and customer satisfaction using survey, and establish a mathematical model between demand and satisfaction [30].

2.2. Online Reviews and Hotel Customer Satisfaction

Identifying the causes of customer satisfaction and dissatisfaction has been an important area of research for researchers from multiple disciplines, including marketing and the hotel industry [31]. In the hotel environment, hotel customers usually assess satisfaction based on their characteristics and services of the hotel. In other words, these characteristics of hotels can not only reveal customers' preferences about hotel characteristics but also characterize their satisfaction. Online reviews and customer ratings of hotels appear to be an excellent tool, as they provide a new avenue to investigate customer satisfaction. Online reviews and review ratings provide an opportunity to observe the causes of customer satisfaction [32–34]. Positive reviews indicate customer satisfaction, while negative reviews indicate customer dissatisfaction [35,36]. Customer satisfaction arises from the comparison between the a priori expectation and the a posteriori perceived performance of a product or service [37]. If consumers perceive higher than expected performance, satisfaction is revealed. In contrast, lower than expected product or service performance can lead to

negative dissatisfaction and dissatisfaction. Hotels are typical service enterprises. Thus, satisfaction is not only the intuitive satisfaction of consumers with regard to product quality, price, service, etc., and a deeper level should be the degree of service provided by hotels or the fit of the products with consumer expectations. Therefore, we consider customer satisfaction as an overall emotional reaction of consumers to the entire set of tangible facilities and intangible services.

Online reviews published by customers best reflect such emotional reactions and are the most direct expression of whether or not the customer is satisfied. One study showed that 73% of respondents preferred to read online consumer reviews about hotels rather than relying on hotels' descriptions of themselves [38]. Hundreds of millions of potential hotel customers visit such online reviews each year [39]. Existing evidence suggests that online users are affected by online reviews in their purchase decisions [40]. Therefore, online reviews have become a relevant source of information for understanding hotel customers. Customers consider online reviews as a reliable source of information when selecting hotels [41,42]. Interestingly, research on online reviews in Tripadvisor showed that online reviews are more important than traditional ways in explaining hotel customer satisfaction. Ye et al. conducted an empirical study using data from a large online travel service in China to determine the impact of online user generated reviews on room sales and business performance in hotels [43]. Vermeulen and Seegers applied the consideration set theory to simulate the impact of online hotel reviews on consumer choice and found that review price, hotel familiarity, and reviewer expertise affect consumer perceptions of hotels [44].

Online reviews have proven to be a rich source of information for analyzing customers' opinions, practices, and behaviors [45,46]. In the hotel industry, the promotion of online reviews has been found to increase customer awareness and scrutiny of the hotel [44,47,48]. Online travel reviews are becoming an influential source of information that affects customers' prepurchase evaluation of hotels. The core parts of online reviews are digital ratings and text reviews [49]. Another study emphasizes that the usefulness of reviews and recommendation centered prices are important factors affecting consumer behavior and purchase decisions [50]. The importance of online review has been widely recognized in the existing literature [51,52]. The existing research to a large extent solved the impact of various rating values on consumer preferences and behavior [53,54]. In addition, the decision-making process of customers is strongly influenced by online reviews [55]. Therefore, it is very important to study the reviews of online hotel recommendation websites to determine customers' behaviors and preferences, because it is an excellent tool to understand customer satisfaction [46,56]. In addition, one of the main advantages of studying customer reviews and ratings is that it can directly show customer satisfaction [32].

2.3. Text Analysis of Online Reviews

Compared with questionnaires, online text reviews are unstructured user generated content [57] that can reflect customer consumption experiences and perceptions in greater detail [35]. Previous studies of online reviews of hotels can be divided into two types.

(1) Some focus on the content of text reviews to find the attributes mentioned by hotel customers and their perceptions of the hotel accommodation experience. These attributes include room quality, staff attitudes and behaviors, location, access, values, food, etc. [35]. Experiences include customer satisfaction and dissatisfaction [33], with positive evaluations indicating customer satisfaction and negative evaluations indicating customer dissatisfaction. Many studies use text mining techniques, including content analysis [58], frequency analysis [59], text linkage analysis [33], and latent semantic analysis [35], which are used to identify attributes of hotel facilities and services that are of customer concern.

(2) The technical aspects of processing online text reviews about hotels have attracted more attention in the research field of online reviews. The big data era has accelerated the use of text mining techniques, which include statistical and machine learning approaches to inform decisions [60]. Text mining is the process of extracting and analyzing large amounts of data from different rules and scenarios to discover hidden patterns in a given

dataset [61], which helps reveal hidden trends regarding products, customers, market trends, and other key factors for the success of the company [62]. Text mining analysis of online reviews can help hotels make predictions, such as review usefulness [63], customer satisfaction [64], and hotel performance [65]. Regarding the relationship between hotel text reviews and customer satisfaction, Geetha et al. focused on the emotional polarity of online customer reviews and find that it affects customer satisfaction [16]. He et al. used natural language pre-processing, text mining, and sentiment analysis techniques to analyze online hotel text reviews, and they found that the affective scores of the titles and contents of online customer reviews are highly correlated with the overall customer ratings of hotels [66]. Qu et al. also found that most attribute emotions derived from text reviews were significantly related to customers' overall ratings [67]. This also supports the findings of Kim et al., who showed that overall rating is the key predictor of hotel performance [68].

Since Chinese consumers are considered typical collectivists and their behaviors are different from those in western countries, what is the impact of online textual reviews of hotels on their satisfaction? Our study helps to address this question. We study the technical aspects of hotel text reviews. First, through text mining of online reviews, we classify and cluster the reviews to obtain the first-level influencing factors on customer satisfaction and conclude the second-level influencing factors on customer satisfaction by referring to the CCSI. Second, we construct the BP neural network and the weight matrix of the hotel customer satisfaction influencing factors model. Here, we take the second-level influencing factors as variables and utilize backpropagation (BP) neural network to do regression analysis on the overall customer satisfaction and calculate the effect size of the first-level influencing factors through the weight matrix of BP neural network.

Finally, according to the intensity of the role of first-level influencing factors, the facilities and services of hotels are adjusted to meet the needs of customers and improve the competitiveness of hotels.

3. Model and Experimental Analysis

This study includes the acquisition and pre-processing of text data, TF-IDF, word2vec, and K-means, and a BP neural network-based regression model of hotel customer satisfaction is constructed by combining CCSI to obtain the effect size difference of each influencing factor on hotel customer satisfaction. The online review text data of hotels were pre-processed, and the TF-IDF value of each word was calculated. This paper defines the 60 top ranked terms of TF-IDF values as the second-level influencing factors on customer satisfaction in hotels. K-means clustering was performed on the second-level influencing factors, clusters of points near the same centroid obtained from clustering were assigned to the same class, and factors in the same class were named first-level influencing factors in combination with CCSI. Through the weight matrix calculation of the BP neural network is able to clearly analyze the customer's characteristic attention preferences to hotels and their characteristic emotions on the satisfaction of hotels, thus helping hotel operators adjust their operation model of hotels to provide better services to customers. The research framework of this paper is shown in Figure 1.

3.1. Data Sources and Pre-Processing

3.1.1. Data Sources

This paper refers to the 2019–2020 China Travel Word of Mouth List released by Ctrip.com, which is inductive based on data such as destination customer voting, searching, and browsing destination content products to select six target cities. Hotel choice is based on: (1) the criteria for selecting star hotels in the studies by Xiang et al. [69] and Geetha et al. [16]; (2) the global hotel leader in 2019 issued by authoritative media "Hotels" in the global hotel industry; and (3) the audience of hotel brands in China. Ten hotel brands (covering 38 sub brands) were selected. Meanwhile, the time frame was limited to one year considering the review timeliness. We obtained 37,146 reviews of the Ctrip.com,

which spans from November 2019 to November 2020 and involves six cities and 122 hotels. Specific information fields for reviews and ratings are shown in Figure 2.

In addition, three-star hotels and above were selected as the research object. From the perspective of hotels, the more comprehensive the star service is, the more emphasis is placed on enhancing the customer experience, thus the review content is very important for hotels to improve the service quality. From the customer perspective, star rated hotels are selected to pursue better services and experiences, and reviews are the best way for them to obtain information and give feedback.
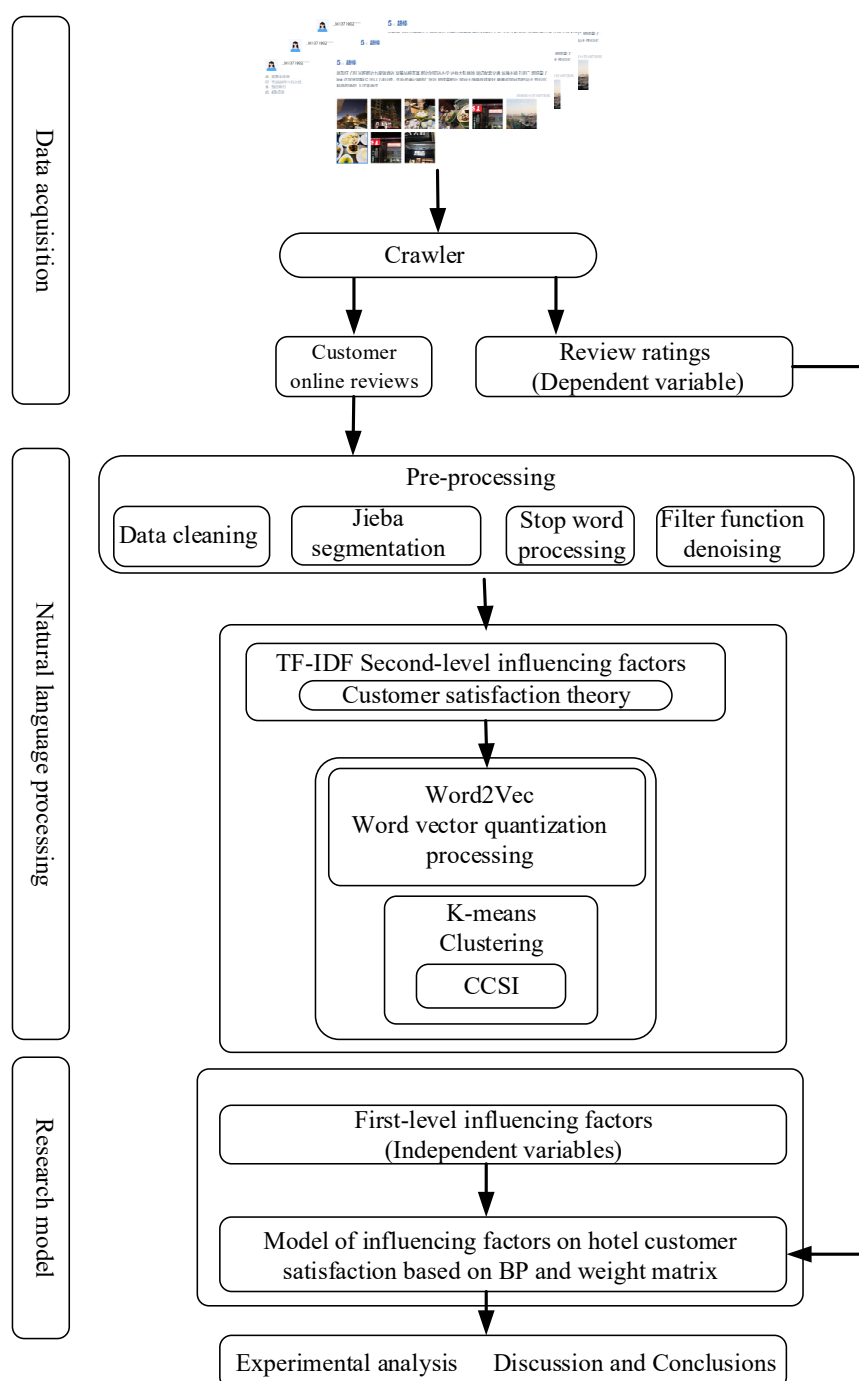

Figure 1. Research framework.

Figure 2. Hotel online reviews and ratings in Ctrip.com.

3.1.2. Data Cleaning

First, we excluded the invalid reviews. The principles are as follows: reviews that are too short or have a few words, reviews with garbled code, and reviews with no analytical significance, which means that there are only symbols or expressions but no words, were removed. Then, we added the word segmentation and stop word library to segment and delete the stop words. Finally, 36,936 valid reviews were obtained after filtering the noise word. The city distribution of the reviews is shown in Table 1.

Table 1. City Distribution of reviews.

Cities	Reviews	Cities	Reviews
Sanya	5952	Xiamen	6049
Shanghai	6566	Canton	6054
Beijing	6341	Hangzhou	5974

Step 1: Jieba word segmentation

Jieba word segmentation has a good segmentation effect for data in the hotel field, and the module design of Jieba segmentation is easy to use and integrate in a system, which can save unnecessary time without affecting the accuracy of segmentation [70]. Therefore, this paper selects Jieba word segmentation tool to process the cleaned data and uses Python Jieba word segmentation interface to segment the text data.

Step 2: Stop word processing

In this paper, we choose the stop list of Harbin Institute of Technology and remove the modal particles in the text, as well as conjunctions such as "and" and "but" in sentences. Combined with the research data of this paper, the words of place names are added to the stop list to filter out the stop words, such as "Hotel", "Sanya", "Shanghai", "Beijing", "Xiamen", "Guangzhou", "Hangzhou", etc. and further remove irrelevant words in the text.

Step 3: Filter the noise words

After the word segmentation results are processed with stop words, the Sklearn filter function is loaded to process the noise data irrelevant to the topic in the text, and the date, non-Chinese words, and numbers in the comments are filtered to obtain the preliminary data processing results.

3.2. Determine the Influencing Factors of Hotel Customer Satisfaction

(1) The TF-IDF algorithm combined with the related theory to hotel customer satisfaction is used to obtain the second-level influencing factors of hotel customer satisfaction. The TF-IDF algorithm meets the need of keyword extraction, and the more important a

word is to the text, the more likely it is to be the keyword of the text [71,72]. The TF-IDF algorithm is calculated according to Equation (1):

$$tf \times idf(i,j) = tf_{ij} \times idf_i = \frac{n_{ij}}{\sum_k n_{kj}} \times \log\left(\frac{|D|}{1+|D_i|}\right) \tag{1}$$

where n_{ij} means how often word i appears in text j; $|D|$ is the total number of texts in the text set; $|D_i|$ is the number of documents in which word i appears in the text set; and the denominator plus 1 is conducted by adopting the Laplacian smoothing, which avoids the occurrence of parts of new words that did not appear in the corpus resulting in a denominator of zero, enhancing the robustness of the algorithm.

After text pre-processing, TF-IDF values for each word are taken in the next step. The pre-existing corpus data are first trained and then loaded with TF-IDF algorithm to train on corpus data [73]. In the process of training, the method of word filtering retains only noun as the key word, combined with hotel customer satisfaction related theory and research, to obtain 60 secondary influencing factors of hotel customer satisfaction, and the experimental results are shown in Table 2.

Table 2. TF-IDF values of second-level influencing factors on hotel customer satisfaction.

Vocabulary	TF-IDF	Vocabulary	TF-IDF	Vocabulary	TF-IDF
Front desk	0.2447	Masks	0.0363	Guest rooms	0.0187
Facility	0.1193	Price	0.0349	Hotel facility	0.0177
Performance price ratio	0.0961	Overall	0.0305	Subway entrance	0.0172
Environment	0.0939	Friends	0.0291	Scenic spot	0.017
Enthusiasm	0.0913	Shopping	0.0265	Courtesy	0.0168
Breakfast	0.0843	Big bed	0.0262	Style	0.0163
Epidemic situation	0.0685	Dining room	0.0256	Variety	0.0159
Geographical location	0.0642	Warm	0.0255	Toilet	0.0157
Experience	0.0622	Expectation	0.0253	Slipper	0.0157
Feeling	0.0622	Air conditioning	0.0237	Old	0.0154
Traffic	0.0612	Parent-child	0.0225	Hardware	0.0153
Airport	0.049	Fruit	0.0216	Children	0.0152
Subway station	0.0478	Cleaners	0.0212	Road	0.015
Attitude	0.0449	Family	0.0211	House price	0.0146
Waiters	0.0443	Pedestrian street	0.021	Refrigerator	0.0139
Sound insulation	0.0429	Shopping mall	0.0209	Windows	0.0138
Complete	0.0423	Balcony	0.0209	Play	0.0135
Quiet	0.0417	Bathtub	0.0197	Details	0.0133

(2) The Word2Vec tool was used to vectorize the words of the second-level influencing factors. The Word2Vec model can well utilize the word context information and the internal structure information of sentences, based on the neural network model to map words into a low-dimensional, dense real number vector. To get the word vector of a certain word in a field, we first pre-processed the corpus of the field; then used the API interface of Gensim module to load Word2Vec to train the word vector model, resulting in the word vector of each word; and finally extracted the quantitative representation of 60 characteristic words obtained from the previous step.

(3) Euclidean distances between the word vectors were calculated, K-means clustering algorithm was used to obtain the clustering results of the word vectors, and silhouette coefficient was used to examine the separation of the clusters and the compactness of the

clusters to evaluate the clustering quality. The K value was constantly adjusted for repeated experiments, and it was found that at k = 10, its contour coefficient reached a maximum value of 0.741, indicating that the K value was optimal at this time. The clustering results are plotted in Figure 3, where the centroids are identified in ordinal sequence, the clusters of points assigned near the same centroid are the same class as the first-level influence factors, and the value of the abscissa represents the boundary range determined by the value of the second-level influence factor word vector and is dimensionless. Through the results of text cluster analysis, combined with customer satisfaction related theory and research, the influencing factors of customer satisfaction in hotels are classified into 10 categories: epidemic prevention, consumption emotion, convenience, environment, facility, catering, target group, perceived value, price, and service. Then, another 60 second-level influencing factors are classified under the first-level influencing factors, and the specific results are shown in Table 3.
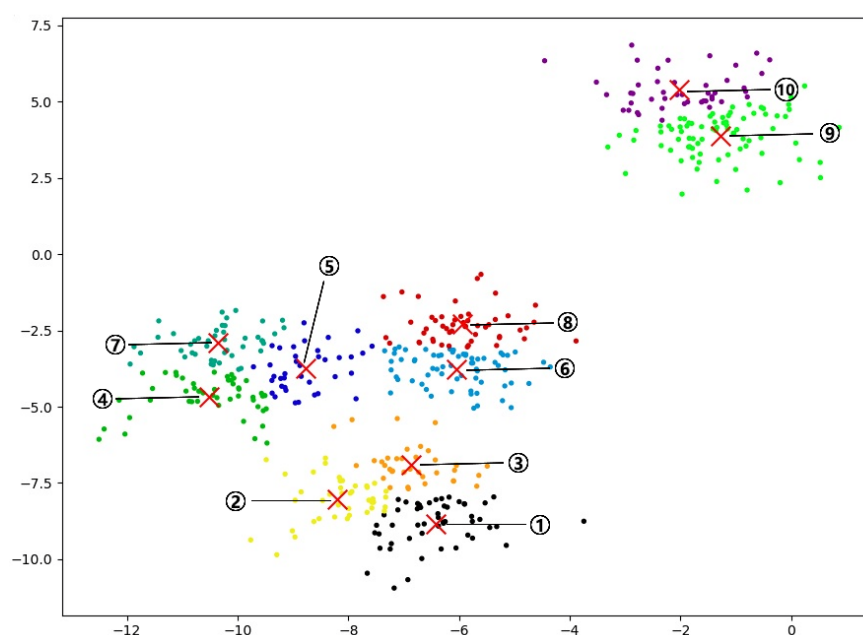

Figure 3. K-means clustering results.

Table 3. Hotel customer satisfaction influencing factors.

First-Level Influencing Factors	Second-Level Influencing Factors
Epidemic prevention	Epidemic situation, Masks
Consumption emotion	Overall, Warm, Attitude, Enthusiasm, Experience, Courtesy, Details, Front desk
Convenience	Airport, Subway entrance, Traffic, Road, Parking lot, Subway station, Geographical location
Environment	Environment, Shopping, Shopping mall, Scenic spot, Play, Pedestrian street
Facility	Air conditioning, Windows, Old, Hotel facility, Facility, Refrigerator, Hardware, Elevator, Sound insulation, Balcony, Bathtub, Toilet, Big bed, Swimming Pool
Catering	Breakfast, Dining room, Fruit, Variety
Target group	Family, Parent-child, Children, Kids, Friends
Perceived value	Performance price ratio, Complete, Quiet, Feeling, Style, Expectation
Price	House price, Price
Service	Waiters, Lobby, Cleaners, Guest rooms, Slipper, Laundry

(4) We determine the value of first-level and second-level influencing factors in each review. First, the value of second-level influencing factors of online reviews of hotels is calculated. The online reviews of hotels are data cleaned, and then each review after cleaning is contrasted with the word of the second-level influencing factors; if one second-level influencing factor is included in the review, the corresponding second-level influencing factors of this review takes the value of 1, and so on, sequentially, adding up the values of all the second-level influencing factors included in the review. Finally, the value of the first-level influencing factor of the hotel's online review takes the value of the corresponding second-level influencing factor additively.

(5) Customer satisfaction rating: On Ctrip.com, the review rating is the customer satisfaction rating of the hotel. Each customer evaluates the hotel as a rating of their satisfaction of the hotel, and the review rating has a value range of 1–5 points, where 5 represents the highest satisfaction. The review rating is directly crawled from the hotel's webpage, and the customer review rating is shown in Figure 2. Customer rating reflects the overall satisfaction during the hotel consumption experience, corresponding to the comments made to the hotels, and, therefore, the rating can well reflect customer satisfaction with the hotels [74].

3.3. Regression Model and Experiment Based on BP Neural Network

The first-level influencing factors of customers' online reviews are taken as independent variables (input) and customer satisfaction scores as dependent variables (output), which are divided into two parts: training set and test set. According to the training set, the customer satisfaction regression model based on BP neural network is constructed by setting and adjusting the corresponding parameters combined with empirical formula. The test set is used to calculate the regression analysis results of the model and verify the trained BP regression model.

3.3.1. Training Set and Test Set

K-fold cross validation method is used to determine the training set and test set, which effectively avoids over learning and under learning [75]. In this paper, K = 10, i.e., 36,936 hotel online reviews are randomly divided into 10 parts based on the satisfaction level, each part is a subset, and 10 different sets of training set and test set are obtained.

3.3.2. BP Network Structure and Experiments

The parameters including the number of nodes in the input layer, the number of hidden layers, the number of nodes in the hidden layer, and the number of nodes in the output layer were determined to set the BP network structure. BP neural network models can contain multiple hidden layers when constructing data relationships, but this does not mean that more hidden layers is always better. Moreover, the junction of the hidden layer in the model is actually the knowledge that the BP model learns through the dataset, and it is able to exhibit complex nonlinear relationships in the data. If the number of hidden layer nodes is too small, the model construction will fail. On the contrary, if the number of hidden layer nodes is too large, it will produce the defect of overfitting, which will make the prediction results have too large error. Therefore, to avoid overfitting in training as much as possible, sufficiently high network performance and generalization ability are guaranteed. The most basic principle for determining the number of hidden layer nodes is to take the most compact structure possible, i.e., to take the smallest possible number of hidden layer nodes, while meeting the accuracy requirements. The appropriate number of nodes was chosen to maximize the performance of the BP neural network. The empirical formula is adopted to select the number of implicit layers and nodes to determine their approximate range, which avoids blinding and improves the effectiveness of the selection [76]. In this paper, the number of hidden layer nodes is determined with reference to empirical Equation (2).

$$S = \sqrt{mn} + a \tag{2}$$

where S is the number of nodes of the hidden layer, *m* is the number of nodes of the input layer, and *n* is the number of nodes of the output layer. *a* is a constant belonging to 1–10.

According to the data features of this paper, the input layer corresponds to the first-level influencing factors, which are epidemiological prevention, consumption emotion, prevention, environment, facility, catalog, target group, administered value, price, and service. Thus, the number of input layer nodes: M = 10. The output layer corresponds to the review rating with the number of nodes in the output layer: n = 1. Drawing from empirical Equation (3) and after several trials in this study, we found that, when the number of hidden layers is 1 and the number of nodes of the implicit layer is 9, the BP neural network has the smallest mean square error (MSE) and the fitting coefficient R is closest to 1. Therefore, we set the number of layers of the implicit layer to 1 and the number of nodes of the implicit layer to 9. The model of the BP neural network is shown in Figure 4.
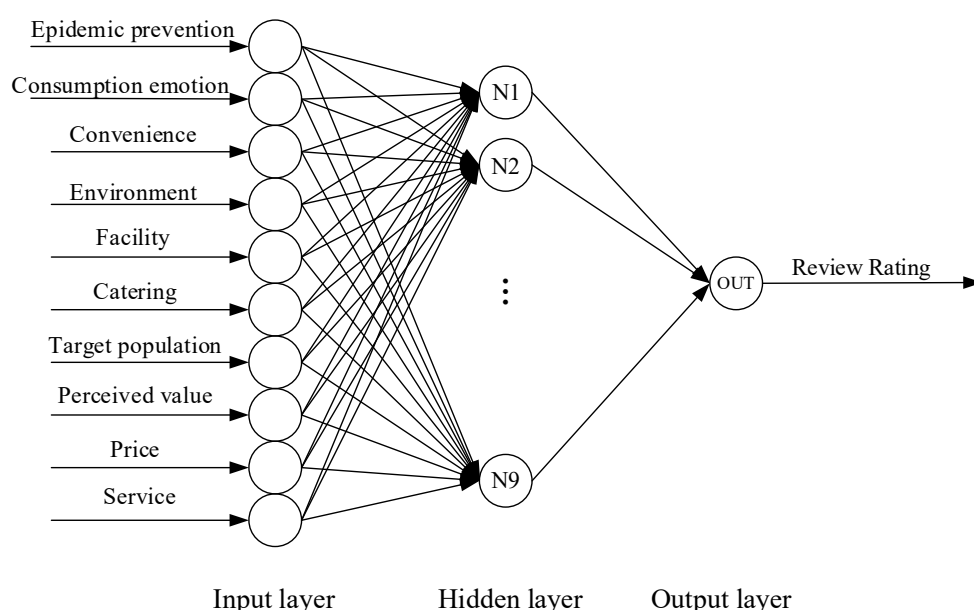

Figure 4. The structure of BP artificial neural network.

In this study, the experimental environment selected MATLAB r2016b neural network toolbox, with the hidden layer node activation function as sigmoid, the output layer neuron activation function as purelin, and the training function as trainlm. Let K = 10, normalize the 36,936 online review data to randomly split into 10 parts, resulting in 10 different sets of training and test sets. The maximum number of training is 10,000 times. The initial value of momentum factor is 0.95, and the initial right value is 0.3.

The basic idea of the BP neural network algorithm is the gradient descent method, which uses the gradient search technique so that the error of the actual output value and the expected output value of the network are the smallest and the best values of input weights and output weights are obtained. After training the neural network, the ideal model is obtained. MSE in the training process is 0.00466, and the network error meets the requirements. The fitting coefficient R = 0.9279, the predicted value is very close to the actual value, and the fitting effect is ideal. To investigate the overall regression prediction ability of BP neural network on customer satisfaction, we calculated the goodness of fit and MSE of the model and compared it with deep belief networks (DBN), support vector machine (SVM), and random forest (RF), as shown in Table 4. The experimental results show that the MSE of BP neural network model is smaller than the other three methods, and the fitting coefficient is the largest. Therefore, BP neural network model is better than DBN, SVM, and RF in the analysis of influencing factors of customer satisfaction.

Table 4. Comparison with other methods.

Model Type	R^2	MSE
BP	0.861	0.00466
DBN	0.855	0.00502
SVM	0.702	0.00621
RF	0.739	0.00602

3.3.3. The Effect of First-Level Factors

We adopt Equation (3) to calculate the relative strength of the first-level influencing factors, i.e., the effect size of the input variables (first-level influencing factors) on the output variable (customer satisfaction) [19,77,78].

$$R_{ji} = \frac{\sum_{k=0}^{n} \left(W_{ki} * W_{jk} \right)}{\sum_{i=0}^{m} | \sum_{k=0}^{n} \left(W_{ki} * W_{jk} \right)} \tag{3}$$

where W_{ki} indicates the weight between the kth hidden unit and the ith input unit. W_{ki} forms the input weight matrices. W_{jk} denotes the weight between the jth output unit and the kth hidden unit. W_{jk} formed the output weight matrices. R_{ji} is the relative strength between the ith input and the jth output variable. R_{ji} is a statistic whose value is the ratio of the strength of the relationship between the ith input variable and the jth output variable to the total strength of all input and output variables.

To estimate the right value of BP neural network, Table 5 presents the input layer to hidden layer right value Wki estimation table and Table 6 presents the hidden layer to output layer right value Wki estimation table. N1–N9 in Tables 5 and 6 indicate the nine nodes of the hidden layer. R_{ji} was calculated by Equation (3). R_{ji} effectively accounts for the relative strength between input and output variables in BP neural network model. Thus, the strength of the effect of first-level influencing factors in online reviews on customer satisfaction is calculated as shown in Table 7.

Table 5. Weights W_{jk} in the path from the input layer to the hidden layer.

Input Layer	N1	N2	N3	Hidden N4	Layer N5	N6	N7	N8	N9
Epidemic prevention	−0.18	0.19	−0.34	0.32	0.36	−0.02	−0.05	0.43	0.17
Consumption emotion	0.76	0.34	−0.06	0.30	0.64	0.29	−1.13	−0.40	0.81
Convenience	0.30	0.48	−0.55	0.22	0.19	−0.27	−0.20	0.51	0.18
Environment	−0.42	0.37	−0.25	−0.38	−0.10	−0.26	0.73	0.09	−0.45
Facility	0.61	0.27	−0.22	−0.50	0.48	−0.11	0.14	−0.01	−0.07
Catering	−0.23	0.17	0.33	−0.60	0.72	−0.14	−0.35	−0.31	−0.19
Target group	−0.70	−0.03	−0.26	0.27	−0.20	0.62	−0.20	0.29	1.02
Perceived value	0.08	−0.73	0.79	0.82	−0.22	0.77	−0.18	0.64	0.31
price	−0.38	−0.47	0.56	0.73	0.40	−0.05	−0.87	−0.28	0.58
service	0.92	−0.09	0.31	−0.71	−0.81	0.52	1.19	−0.10	−0.81

Table 6. Weights W_{jk} in the path from the hidden layer to the output layer.

Output Layer	Hidden Layer								
	N1	N2	N3	N4	N5	N6	N7	N8	N9
Review Rating	−0.43	−0.22	0.55	0.35	−0.11	0.56	0.46	−0.12	−0.27

Table 7. Relative strength values of the effect of first-level factors on customer satisfaction.

First-Level Influencing Factors	Relative Strength (R_{ji})
Epidemic prevention	0.1071
Consumption emotion	0.1741
Convenience	0.0954
Environment	−0.0751
Facility	0.0592
Catering	−0.0740
Target group	0.0934
Perceived value	0.2397
Price	0.0248
Service	0.0472

Through the above experimental analysis, epidemic prevention, consumption emotion, convenience, catering, target group, perceived value, price, service, and customer satisfaction are positively related, while environment, facility, and customer satisfaction are negatively related. The relative strengths of consumption emotion, perceived value, epidemic prevention, target group, and convenience are greater than 0.06, and they significantly affect customer satisfaction. The relative strength of perceived value is significantly greater than the odds ratios of other factors, indicating that the value that hotels offer to customers has the greatest impact on customer ratings. The second is consumption emotion, whereby customers value the emotional experience that the hotel's services bring to them. It is worth noting that epidemic prevention emerged as a new factor affecting customer satisfaction, which has not appeared in previous studies, but is in line with the concern of customers traveling for their own health and residential environment under the current outbreak. The environment, facility, catering, and service have less impact on customer satisfaction, while price has the least impact. This may be because customers who choose star rated hotels have stronger affordability and the price has some effect on their satisfaction, but only a few customers are conscious to express concern about price in the reviews.

4. Discussion and Conclusions

4.1. Conclusions

This study analyzed 36,936 hotel online reviews by text mining and summarized 10 first-order factors affecting hotel customer satisfaction based on clustering results and the CCSI: epidemic prevention, consumption emotion, convenience, environment, facility, catering, target group, perceived value, price, and service. A regression model combining BP neural network and weight matrix was constructed to analyze the relationship between first-level influencing factors and customer satisfaction to provide practical suggestions for enhancing customer satisfaction.

With the above experimental analysis, we have some interesting findings. Through text mining of online reviews of hotels, we obtained some new second-level influencing factors, which are shown in Table 3. Combining the CCSI yields some new first-level influencing factors on customer satisfaction of hotels, which are shown in Table 3. Modeling the BP neural network and the right matrix, we obtained the relative strength of the first-level influencing factors (Table 7), and the results show that some influencing factors' effects are not consistent with the existing studies. This paper emphasizes the role of consumption emotion, perceived value, epidemic prevention, target group, and convenience in customer perception and satisfaction of hotels. which also further supports the conclusion of Wu et al. [79], who incorporated perceived value into customer satisfaction evaluation indicators. It is known from our findings that perceived value directly affects customer satisfaction and plays an important role. At the same time, we also found some factors that rarely appear in previous studies on hotel satisfaction, such as epidemic prevention, target group, and consumption emotion. Epidemic prevention

reflects the current customer concerns about infectious disease prevention, with customers focusing on physical health and self-protection throughout travel. For example, hotels provide disposable medical masks free of charge and bring care and warmth to customers, which becomes an important influencing factor for customer satisfaction. Nilashi classified tourist groups into four categories, namely travelling solo, travelling with family, travelling as a couple, and travelling with friends, and suggested that different needs arise from different groups of customers [80], which is similar to the conclusion of our study. Target group was often mentioned in online reviews, referring to groups traveling for concern in a customer perspective, such as families, parent–child, children, and friends. This suggests that customers generally enjoy travel times with loved ones and are more concerned about the feelings and experiences of their peers than with themselves. Consumption emotion refers to a series of emotional reactions that arise during product use and consumption experience. In the whole consumption process, all the links affect customers' consumption emotion. The good and bad emotional experiences of customers directly determine the high and low levels of satisfaction. Convenience is also an important factor influencing customer satisfaction, where airport and subway have become the keys for people to choose hotels. Airport helps customers arrive at their destination city quickly, and subway helps customers travel through the city for easy congestion. Price is a factor that often emerges in satisfaction research [81,82]. In previous studies, price is one of the important factors affecting customer satisfaction and purchase decisions in hotels [3]. However, this study shows that price has little relevance to our systematic objectives, that is, price has less impact on customer satisfaction in hotels. It can be seen that customers who choose hotels with three stars or more have generally high economic strength and are not price sensitive. Besides price, customers value perceived value and consumption experience more. Environment, facility, and catering factors that are frequently observed in previous studies on customer satisfaction are still influencing customers' ratings of the hotel, but they are less influential compared to consumption emotion, perceived value, epidemic prevention, target group, and convenience.

The experiment in this paper shows that environment and catering are negatively related to customer satisfaction, which indicates that customer dissatisfaction is mainly focused on environment and catering. The negative correlation between environment and satisfaction may be that some hotels ignore customers' need for shopping and playing in accommodation. Customers have a strong shopping demand during travel, prefer to take some of the more characteristic goods home, and prefer to buy them at larger shopping malls. Catering is negatively related to satisfaction, and many customers are not satisfied with the breakfast at the hotel, feeling that the type of breakfast is less diverse and lacked fruits.

4.2. Theoretical Significance

First, this paper combines Chinese customer satisfaction index, natural language processing, BP neural network, and weight matrix calculation to construct a new theoretical model to study the influencing factors of hotel customer satisfaction, and then discusses the proposed research questions. The main purpose of this study was to verify and identify the influencing factors that affect customer satisfaction of hotels based on online reviews, as well as to explore the influence and polarity of these influencing factors. The model of hotel customer satisfaction constructed in this paper can be realized to extend to other research areas in the future.

Second, from the perspective of influencing factor acquisition methods, the authentic evaluations of customer expressions in online reviews are extracted using text mining techniques, which are more objective and authentic for data acquisition than traditional questionnaires. Sometimes customers do not fill out questionnaires, but go to their hotel ordering site to share their positive or negative experiences. Therefore, finding the feelings and experiences of hotels in online reviews of customer hotels is the premise and key to research on customer satisfaction.

Third, from a methodological perspective, the BP neural network and the weight matrix can be used in the study of hotel customer satisfaction. BP neural networks employ a data-driven adaptive approach that requires only few a priori assumptions about the model. Meanwhile, the application of the right matrix calculation makes the neural network more suitable than the linear model to describe the relationship between complex variables. The BP neural network and the weight matrix in this study can help hotels discover the decisive factors of customer satisfaction by analyzing a large number of online reviews. Therefore, this paper combines text mining, BP neural network and weight matrix to study hotel customer satisfaction, which is a meaningful attempt.

4.3. Management Implications

Review rating is a direct measure of customer satisfaction. Text reviews are indirect measures of customer perceptions and satisfaction that adequately reflect customer consumption experiences. The results of this study can motivate hotels to mine more attributes from customer text reviews and deeply investigate the relationship between online reviews and customer satisfaction. These online reviews produced a high word-of-mouth effect that influenced reservation decisions of future customers. This study analyzed the factors that affect customer satisfaction of hotels in online reviews, with some practical management implications.

First, new needs of customers are identified from online reviews in a timely manner. Crawler is employed to obtain online reviews of customers, which is more real-time and excavates new factors affecting customers through natural language processing. Then, BP neural network and weight matrix are adopted to calculate the accurate analysis of the influence size and polarity of these new factors. This study shows that pandemic prevention factors become new factors affecting customer satisfaction, and the relative impact strength of pandemic prevention is even higher than convenience. In the period of COVID-19, pandemic prevention is a problem for everyone. Hotels need to provide measures against pandemic to meet customers' new needs, for example, providing disposable medical masks free of charge every day during the hotel stay, and disinfecting rooms and public areas daily.

Second, hotels should accurately implement online marketing. The results of this study indicate that the high relative impact strength of the target group is one of the key factors affecting consumer satisfaction. Different types of customers have different consumption preferences and target customers can be accurately located through online reviews. Therefore, hotel managers should take different steps to provide differentiated services to maximize the demand for different customers. According to changes such as time and holidays, combining demand differences of different groups and consumption pain points provides a differentiated portfolio of products, e.g., introducing parent–child rooms, couple rooms, business rooms, etc.

Third, hotels must keep up with customer consumption changes. This study showed that customers of Hotels with three or more stars are insensitive to price factors, which means that managers cannot rely excessively on price to mobilize consumer perceived value, but instead they should focus more time and effort on other factors to enhance customer satisfaction by improving service quality, facility quality, and creating a more convenient environment.

Fourth, hotels are to profit in cooperation. Hotels should work with their surrounding neighbors, as this study showed that malls, walking streets, and the quantity and quality of attractions are all important external factors in a hotel. Customers, in addition to accommodation needs, also have the need to shop and play during travel. Hotels can cooperate with local malls or specialty shops to launch the hotel's memorial set of goods. Hotels can also cooperate with scenic spots and open direct free shuttle bus from hotels to scenic spots.

5. Future Research Directions

The limitations of this paper are its research scope and general applicability. The current study data come from hotel online reviews in six cities on carrier ship online, and whether these results are applicable to other regions and web ordering platforms remains to be verified. Whether the conclusion that price is not a major factor affecting customer satisfaction is applicable to hotels in other regions remains to be examined. Future research can further investigate using hotel reviews from other regions and web ordering platforms as research objects to take review data from other websites (e.g., www.booking.com and www.tripadvisor.cn).

Author Contributions: Conceptualization, J.W. and Z.Z.; methodology, J.W.; software, Y.G.; validation, J.W. and Y.L.; formal analysis, J.W.; investigation, Y.L.; resources, J.W.; data curation, Y.L.; writing—original draft preparation, J.W.; writing—review and editing, Z.Z.; and visualization, Y.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1.　Li, H.; Liu, Y.; Tan, C.-W.; Hu, F. Comprehending customer satisfaction with hotels: Data analysis of consumer-generated reviews. *Int. J. Contemp. Hosp. Manag.* **2020**, *32*, 1713–1735. [CrossRef]
2.　Tseng, T.H.; Chang, S.-H.; Wang, Y.-M.; Wang, Y.-S.; Lin, S.-J. An Empirical Investigation of the Longitudinal Effect of Online Consumer Reviews on Hotel Accommodation Performance. *Sustainability* **2021**, *13*, 193. [CrossRef]
3.　Lee, M.; Cai, Y.M.; DeFranco, A.; Lee, J. Exploring influential factors affecting guest satisfaction Big data and business analytics in consumer-generated reviews. *J. Hosp. Tour. Technol.* **2020**, *11*, 137–153.
4.　Duan, W.; Yu, Y.; Cao, Q.; Levy, S. Exploring the impact of social media on hotel service performance: A sentimental analysis approach. *Cornell Hosp. Q.* **2016**, *57*, 282–296. [CrossRef]
5.　Xiong, W.; Xu, J. A evaluation on service quality of economic hotel in China: A content analysis of guest comments on website. *J. Beijing Int. Stud. Univ.* **2010**, *11*, 57–67.
6.　Tsaur, S.-H.; Lin, Y.-C. Promoting service quality in tourist hotels: The role of HRM practices and service behavior. *Tour. Manag.* **2004**, *25*, 471–481. [CrossRef]
7.　Pine, R.; Phillips, P. Performance comparisons of hotels in China. *Int. J. Hosp. Manag.* **2005**, *24*, 57–73. [CrossRef]
8.　Rodríguez-Díaz, M.; Rodríguez-Díaz, R.; Rodríguez-Voltes, A.C.; Rodríguez-Voltes, C.I. A model of market positioning of destinations based on online customer reviews of lodgings. *Sustainability* **2018**, *10*, 78. [CrossRef]
9.　Lee, J.; Benjamin, S.; Childs, M. Unpacking the Emotions behind TripAdvisor Travel Reviews: The Case Study of Gatlinburg, Tennessee. *Int. J. Hosp. Tour. Adm.* **2020**, 1–18. [CrossRef]
10.　Liu, J.; Cao, Y.; Lin, C.-Y.; Huang, Y.; Zhou, M. Low-quality product review detection in opinion summarization. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic, 28–30 June 2007; pp. 334–342.
11.　Eirinaki, M.; Pisal, S.; Singh, J. Feature-based opinion mining and ranking. *J. Comput. Syst. Sci.* **2012**, *78*, 1175–1184. [CrossRef]
12.　Abbasi, A.; Chen, H.; Salem, A. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.* **2008**, *26*, 1–34. [CrossRef]
13.　Ganu, G.; Kakodkar, Y.; Marian, A. Improving the quality of predictions using textual information in online user reviews. *Inf. Syst.* **2013**, *38*, 1–15. [CrossRef]
14.　Archak, N.; Ghose, A.; Ipeirotis, P.G.J.M.S. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Manag. Sci.* **2011**, *57*, 1485–1509. [CrossRef]
15.　Ríos-Martín, M.Á.; Folgado-Fernández, J.A.; Palos-Sánchez, P.R.; Castejón-Jiménez, P. The Impact of the Environmental Quality of Online Feedback and Satisfaction When Exploring the Critical Factors for Luxury Hotels. *Sustainability* **2020**, *12*, 299. [CrossRef]
16.　Geetha, M.; Singha, P.; Sinha, S. Relationship between customer sentiment and online customer ratings for hotels—An empirical analysis. *Tour. Manag.* **2017**, *61*, 43–54. [CrossRef]

17. Ban, H.-J.; Choi, H.; Choi, E.-K.; Lee, S.; Kim, H.-S. Investigating key attributes in experience and satisfaction of hotel customer using online review data. *Sustainability* **2019**, *11*, 6570. [CrossRef]

18. Sim, J.J.; Tan, W.H.; Wong, J.C.J.; Ooi, K.B.; Hew, T.S. Understanding and predicting the motivators of mobile music acceptance—A multi-stage MRA-artificial neural network approach. *Telemat. Inform.* **2014**, *31*, 569–584. [CrossRef]

19. Zhao, Z.; Wang, J.; Sun, H.; Liu, Y.; Fan, Z.; Xuan, F. What Factors Influence Online Product Sales? Online Reviews, Review System Curation, Online Promotional Marketing and Seller Guarantees Analysis. *IEEE Access* **2019**, *8*, 3920–3931. [CrossRef]

20. Chong, A.Y.L.; Li, B.; Ngai, E.W.T.; Ch'ng, E.; Lee, F. Predicting online product sales via online reviews, sentiments, and promotion strategies A big data architecture and neural network approach. *Int. J. Oper. Prod. Manag.* **2016**, *36*, 358–383. [CrossRef]

21. George, G.; Haas, M.R.; Pentland, A. Big Data and Management. *Acad. Manag. J.* **2014**, *57*, 321–326. [CrossRef]

22. Yu, Z. Research on the Companies to Understand Consumers' Psychology and Behavior in the Era of Internet of Things. *East China Econ. Manag.* **2012**, *1*, 10.

23. Tontini, G. Integrating the Kano model and QFD for designing new products. *Total Qual. Manag.* **2007**, *18*, 599–612. [CrossRef]

24. Shi, L. Supply chain partnership research based on customer satisfaction degree. In Proceedings of the 2012 Second International Conference on Business Computing and Global Informatization, Shanghai, China, 12–14 October 2012; pp. 203–205.

25. Li, F.; Wei, D. The impact of information and information system satisfaction on customer satisfaction under e-commerce. In Proceedings of the 2011 International Conference on Information Management, Innovation Management and Industrial Engineering, Shenzhen, China, 26–27 November 2011; pp. 190–193.

26. Liu, X.L. Evaluation theory and application of logistics enterprise customer satisfaction degree. In *Advanced Materials Research*; Trans Tech Publications Ltd.: Freienbach, Switzerland, 2013; pp. 1155–1160.

27. Yang, H.; Wang, W. A Study on Customer Satisfaction Assessment System of Third Party Logistics. *Manag. Rev.* **2015**, *27*, 181–193.

28. Schneider, B.; Macey, W.H.; Lee, W.C.; Young, S.A. Organizational service climate drivers of the American Customer Satisfaction Index (ACSI) and financial and market performance. *J. Serv. Res.* **2009**, *12*, 3–14. [CrossRef]

29. Chen, X.; Li, R.; Wu, X.; Chen, K.; Mo, C. Research on Evaluation Methods of Mass Customization-Oriented Customer Demand Satisfaction. *J. Zhejiang Sci Tech Univ.* **2013**, *23*, 02.

30. Sun, Y.; Liu, F.; Li, L. Importance determining method of personalized product attributes based on Kano-QFD integration model. *Comput. Integr. Manuf. Syst.* **2014**, *20*, 2697–2704.

31. Chi, C.G.-Q.; Qu, H. Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. *Tour. Manag.* **2008**, *29*, 624–636. [CrossRef]

32. Zhao, Y.; Xu, X.; Wang, M. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *Int. J. Hosp. Manag.* **2019**, *76*, 111–121. [CrossRef]

33. Berezina, K.; Bilgihan, A.; Cobanoglu, C.; Okumus, F. Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *J. Hosp. Mark. Manag.* **2016**, *25*, 1–24. [CrossRef]

34. 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]

35. Xu, X.; Li, Y. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *Int. J. Hosp. Manag.* **2016**, *55*, 57–69. [CrossRef]

36. Xu, X. Does traveler satisfaction differ in various travel group compositions? *Int. J. Contemp. Hosp. Manag.* **2018**. [CrossRef]

37. Oliver, R.L. A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Mark. Res.* **1980**, *17*, 460–469. [CrossRef]

38. Saura, J.R.; Reyes-Menendez, A.; Alvarez-Alonso, C. Do Online Comments Affect Environmental Management? Identifying Factors Related to Environmental Management and Sustainability of Hotels. *Sustainability* **2018**, *10*, 3016. [CrossRef]

39. Saura, J.R.; Reyes-Menendez, A.; Filipe, F. Comparing Data-Driven Methods for Extracting Knowledge from User Generated Content. *J. Open Innov. Technol. Mark. Complex.* **2019**, *5*, 74. [CrossRef]

40. Mauri, A.G.; Minazzi, R. Web reviews influence on expectations and purchasing intentions of hotel potential customers. *Int. J. Hosp. Manag.* **2013**, *34*, 99–107. [CrossRef]

41. Gavilan, D.; Avello, M.; Martinez-Navarro, G. The influence of online ratings and reviews on hotel booking consideration. *Tour. Manag.* **2018**, *66*, 53–61. [CrossRef]

42. Yan, Q.; Zhou, S.; Wu, S. The influences of tourists' emotions on the selection of electronic word of mouth platforms. *Tour. Manag.* **2018**, *66*, 348–363. [CrossRef]

43. Ye, Q.; Law, R.; Gu, B.; Chen, W. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Comput. Hum. Behav.* **2011**, *27*, 634–639. [CrossRef]

44. Vermeulen, I.E.; Seegers, D. Tried and tested: The impact of online hotel reviews on consumer consideration. *Tour. Manag.* **2009**, *30*, 123–127. [CrossRef]

45. Költringer, C.; Dickinger, A. Analyzing destination branding and image from online sources: A web content mining approach. *J. Bus. Res.* **2015**, *68*, 1836–1843. [CrossRef]

46. Liu, Y.; Teichert, T.; Rossi, M.; Li, H.; Hu, F. Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tour. Manag.* **2017**, *59*, 554–563. [CrossRef]

47. Horner, S.; Swarbrooke, J. *Consumer Behaviour in Tourism*; Routledge: Abingdon, UK, 2016.

48. Xiang, Z.; Gretzel, U. Role of social media in online travel information search. *Tour. Manag.* **2010**, *31*, 179–188. [CrossRef]

49. Fang, B.; Ye, Q.; Kucukusta, D.; Law, R. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tour. Manag.* **2016**, *52*, 498–506. [CrossRef]

50. Ghosh, T. Predicting hotel book intention: The influential role of helpfulness and advocacy of online reviews. *J. Hosp. Mark. Manag.* **2018**, *27*, 299–322. [CrossRef]

51. Hennig-Thurau, T.; Gwinner, K.P.; Walsh, G.; Gremler, D.D. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *J. Interact. Mark.* **2004**, *18*, 38–52. [CrossRef]

52. Litvin, S.W.; Goldsmith, R.E.; Pan, B. Electronic word-of-mouth in hospitality and tourism management. *Tour. Manag.* **2008**, *29*, 458–468. [CrossRef]

53. Banerjee, S.; Chua, A.Y. In search of patterns among travellers' hotel ratings in TripAdvisor. *Tour. Manag.* **2016**, *53*, 125–131. [CrossRef]

54. Camilleri, A.R. The Presentation Format of Review Score Information Influences Consumer Preferences Through the Attribution of Outlier Reviews. *J. Interact. Mark.* **2017**, *39*, 1–14. [CrossRef]

55. Goldenberg, J.; Libai, B.; Muller, E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Mark. Lett.* **2001**, *12*, 211–223. [CrossRef]

56. Gao, B.; Li, X.; Liu, S.; Fang, D. How power distance affects online hotel ratings: The positive moderating roles of hotel chain and reviewers' travel experience. *Tour. Manag.* **2018**, *65*, 176–186. [CrossRef]

57. Zhang, X.; Yu, Y.; Li, H.; Lin, Z. Sentimental interplay between structured and unstructured user-generated contents: An empirical study on online hotel reviews. *Online Inf. Rev.* **2016**, *40*, 119–145. [CrossRef]

58. Li, H.; Ye, Q.; Law, R. Determinants of customer satisfaction in the hotel industry: An application of online review analysis. *Asia Pac. J. Tour. Res.* **2013**, *18*, 784–802. [CrossRef]

59. Xiang, Z.; Schwartz, Z.; Gerdes, J.H., Jr.; Uysal, M. What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int. J. Hosp. Manag.* **2015**, *44*, 120–130. [CrossRef]

60. Shmueli, G.; Bruce, P.C.; Yahav, I.; Patel, N.R.; Lichtendahl, K.C., Jr. *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*; John Wiley & Sons: Hoboken, NJ, USA, 2017.

61. Maheshwari, A. *Business Intelligence and Data Mining*; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2014.

62. Chu, W.W. Data Mining and Knowledge Discovery for Big Data. *Stud. Big Data* **2014**, *2013*, 601–622.

63. Ma, Y.; Xiang, Z.; Du, Q.; Fan, W. Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *Int. J. Hosp. Manag.* **2018**, *71*, 120–131. [CrossRef]

64. Ludwig, S.; Ruyter, K.D.; Friedman, M.; Brüggen, E.C.; Pfann, G. More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *J. Mark.* **2013**, *77*, 87–103. [CrossRef]

65. Blal, I.; Sturman, M.C. The Differential Effects of the Quality and Quantity of Online Reviews on Hotel Room Sales. *Cornell Hosp. Q.* **2014**, *55*, 365–375. [CrossRef]

66. He, W.; Tian, X.; Tao, R.; Zhang, W.; Yan, G.; Akula, V. Application of social media analytics: A case of analyzing online hotel reviews. *Online Inf. Rev.* **2017**, *41*, 921–935. [CrossRef]

67. Qu, Z.; Zhang, H.; Li, H. Determinants of online merchant rating: Content analysis of consumer comments about Yahoo merchants. *Decis. Support Syst.* **2008**, *46*, 440–449. [CrossRef]

68. Kim, W.G.; Lim, H.; Brymer, R.A. The effectiveness of managing social media on hotel performance. *Int. J. Hosp. Manag.* **2015**, *44*, 165–171. [CrossRef]

69. 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.* **2017**, *58*, 51–65. [CrossRef]

70. Tu, M. *Core Technology and Algorithm of Python Natural Language Processing*; Machinery Industry Press: South Norwalk, CT, USA, 2018.

71. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning, Bejing, China, 22–24 June 2014; pp. 1188–1196.

72. Vrbanec, T.; Meštrović, A. Corpus-Based Paraphrase Detection Experiments and Review. *Information* **2020**, *11*, 241. [CrossRef]

73. Liu, Y.; Zhang, Y.; Zhao, Z.; Yan, X.; Zhang, J. Hotel customer satisfaction evaluation based on sentiment analysis. *Comput. Appl. Softw.* **2020**, *37*, 54–67.

74. Mellinas, J.P.; Nicolau, J.L.; Park, S. Inconsistent behavior in online consumer reviews: The effects of hotel attribute ratings on location. *Tour. Manag.* **2019**, *71*, 421–427. [CrossRef]

75. Hu, J.; Zhang, G. K-fold cross-validation based selected ensemble classification algorithm. *Bull. Sci. Technol* **2013**, *29*, 115–117.

76. Yang, X. Study of early warning for cerebrovascular risk based on deep beliefs networks. *Beijing Jiaotong Univ.* **2016**, *26*, 19–26.

77. Lee, S.; Choeh, J.Y. Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Syst. Appl.* **2014**, *41*, 3041–3046. [CrossRef]

78. Yoon, Y.; Guimaraes, T.; Swales, G. *Integrating Artificial Neural Networks with Rule-Based Expert Systems*; Elsevier Science Publishers B. V.: Amsterdam, The Netherlands, 1994; pp. 497–507.

79. Wu, H.-C.; Ai, C.-H.; Yang, L.-J.; Li, T. A study of revisit intentions, customer satisfaction, corporate image, emotions and service quality in the hot spring industry. *J. China Tour. Res.* **2015**, *11*, 371–401. [CrossRef]

80. 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* **2019**, *11*, 6013. [CrossRef]
81. Schuckert, M.; Liang, S.; Law, R.; Sun, W. How do domestic and international high-end hotel brands receive and manage customer feedback? *Int. J. Hosp. Manag.* **2019**, *77*, 528–537. [CrossRef]
82. Ren, L.; Qiu, H.; Wang, P.; Lin, P.M. Exploring customer experience with budget hotels: Dimensionality and satisfaction. *Int. J. Hosp. Manag.* **2016**, *52*, 13–23. [CrossRef]