Understanding customer experience using online reviews in the hotel industry with a text mining approach (Five-star hotels in Iran)

Zohre Anisi Hemaseh[a], Mohammad Ali Afsharkazemi[b,*], Ehsan Mousavi Khanghah[c]

[a]Department of Information Technology Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran

[b]Department of Industrial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran

[c]Department of Computer Engineering, Shahed University, Tehran, Iran

(Communicated by Mohammad Rasoul Velayati)

Abstract

The tourism industry, with its strategic role in the economic prosperity of a country, has become a passage for sustainable development. One of the necessities of every business is to know the preferences and understand the customer experience. Understanding the customer experience is often done through questionnaires and surveys, which have limitations that have caused the answers to be inaccurate, as well as the orientation of the questions based on the researcher's mind. With the emergence of social networks, users present their opinions regarding the experience they have with products and services. This causes the production of valuable big data. Online reviews are one such source of data, created from customers' self-reports of their experiences. Considering that the hotel industry in Iran needs to develop and increase its share of the international market, the online reviews of five-star hotels in Iran were analyzed on the world-renowned platform. In this study, the approach of frequency analysis and semantic network analysis was used to extract topics in the perceived experience of customers, and then using factor analysis, hidden factors were discovered. Finally, to model the relationship between the factors and their effect on satisfaction, a linear regression model was performed. This research obtained valuable findings from the opinions of customers that in addition to the major role of staff behaviour, room facilities in satisfaction, the different roles of food service in different meals, the effect of the beauty of the city and the purpose of the trip can be mentioned.

Keywords: tourism, online reviews, customer experience, customer satisfaction, semantic network analysis
20220 MSC: 91D30

1 Introduction

In recent years, the tourism industry, as one of the largest industries in the world, has had a strategic role in creating job opportunities and generating currency [5, 24], this industry is a gateway for sustainable development [34]. In Iran, due to the historical and natural attractions, the importance of paying attention to the development of the

---

*Corresponding author

Email addresses: zohre.anissi@gmail.com (Zohre Anisi Hemaseh), dr.mafshar@gmail.com (Mohammad Ali Afsharkazemi), emousavi@shahed.ac.ir (Ehsan Mousavi Khanghah)

tourism industry is growing [29] In addition to the industries related to the tourism industry, hotel management is of special importance as one of the main prerequisites of the tourism industry. Therefore, in the tourism industry, a large investment is made every year for the development of the hotel industry to improve the quality of services to increase customer satisfaction.

The vital success of this industry is possible by forming loyalty from customers and creating a better experience for them. Knowing the factors affecting customer satisfaction and dissatisfaction is a turning point that reveals the hidden expectations and needs of the customer [5]. Among other important issues that organizations deal with in the current competitive environment, are the awareness of customer satisfaction and their views on the performance of the organization, products and services [5] Managers must have a deep understanding of what their customers want. and assign the most value to meet their needs [16] With the emergence of social networks and big data production, a special ability to hear the customer's voice about their needs and preferences emerged, which can actually be considered as a strategic resource [15], key and value-creating [28].

The hotel industry has the opportunity to gain an understanding of customers' experiences by using online comments to identify existing strengths and weaknesses [28] The intangible characteristics of products and services in this field have completely complicated consumer behaviour patterns. Therefore, understanding the customer experience is important to maintain customer satisfaction and lasting loyalty [3], hence the necessity of using the results of online opinion analysis and using this information as a valuable source of customer experience and preferences in the decision-making process of managers [20] is known to improve its marketing strategy and decision-making [1] The questions that will be answered in this study are:

1- How can you get an understanding of customer experience using online reviews?

2- What are the preferences of customers (the strengths and weaknesses of hotels from the customer's point of view) in using five-star hotels?

3- What is the relationship between customer preferences and their satisfaction at the level of five-star hotels?

To understand the status of hotels, in this study, an international approach has been used to collect big data available on the internet and international platforms. These platforms have the ability so that people can share their opinions about any hotel from any place and other people can easily access these experiences. One of these popular platforms is TripAdvisor, where user-generated data can provide significant useful information for managers to provide better services to their customers [20].

This article expands as follows; After the introduction, in the background section of the research, the theoretical issues and studies conducted in the fields of online customer comments, text analysis and semantic networks on big data have been discussed. In the methodology section, the method of data collection and analysis is described with a focus on textual big data. In the results and findings section, frequency analysis, semantic network analysis, and CONvergence of iterated Correlations (CONCOR) were used to extract effective general topics, and then Factor Analysis and Linear Regression were used to determine the relationship between the extracted characteristics and customer satisfaction. has been The last part of the study and the implications of this study for academicians and managers are summarized and the limitations and directions of future research are presented.

2 Research background

2.1 Electronic word of mouth (eWOM) and online reviews

Electronic word-of-mouth is defined as any positive or negative comment made by potential, actual or former customers about a product or service, which is made available to many people and institutions through the Internet [3]. This possibility provides rich data that reflects the perception of satisfaction or dissatisfaction with services and products [1].

Online reviews are an important part of electronic word of mouth [17, 25]. Today, customers mostly tend to present their personal preferences through online opinions [20] to the extent that Fu and Wei [15] stated that these opinions are more valuable than the information provided by businesses. Online reviews contain information about user experience and how that experience is perceived [21], which provides a new perspective for market identification by listening to the voice of customers [21, 31, 32].

2.2 Online reviews, preferences and customer satisfaction

Customer satisfaction is the main indicator of meeting (not) customer needs. According to the expectation disconfirmation theory, when the customer's expectations and perceived services are different, it can affect customer satisfaction [7]. The theory of mismatch of expectations is a very important issue in the service industry where intangible services are provided and satisfaction and revisiting with these services are evaluated [14] The quality of products and services provided affects the expectations of customers [22] According to this theory, Satisfaction is defined as "the result of disconfirmation of performance from expectations". Positive (satisfaction) and negative (dissatisfaction) disconfirmation occur when a provider provides a better/worse experience than expected [23] Therefore, finding customer preferences is part of the effort to improve service quality and Satisfaction, and gaining strategic competitive advantage [15]

Online reviews describe customer experiences and customer satisfaction levels [22]. By collecting and analyzing online opinions, it is possible to understand the feelings, perceptions and results of each person's travel experience, and based on that, provide services that are customized to the characteristics and preferences of users [15] Online reviews are a popular research field for evaluating customer satisfaction [22] and can serve as a strategy for loyalty, relational communication, and creating shared value [27] The process of analyzing online reviews is an important source of information for businesses because it It can help to understand the needs of customers and increase the quality of products/services through marketing strategy [1]. The following table shows the studies conducted on online comments.

Table 1: Previous studies with online opinion analysis approach

Authors	The area under investigation	Application of online comments in the article
Chang Y. et al [7]	air Lines	Subject-based sentiment identification and unrated sentiment prediction
Tao S. and Kim H. [34]	Coffee Shop	Extracting key factors in customer satisfaction
Celuch K [6]	Event management	Extracting effective features and evaluating consumers' feelings towards their experience of buying event tickets
Ding K. et al. [11]	Tourism	Extracting features that affect user satisfaction and dissatisfaction
Fernandes E. et al. [13]	Restaurant management	Analyzing data from different sources and combining them into the perceived value of business information for efficient prioritization of actions
Chatterjee S. and Mandal P. [9]	air Lines	Extracting customer preferences - customer satisfaction analysis
Kitsios F. et al. [23]	hotel management	Extracting and comparing the features of services that are effective in the satisfaction and dissatisfaction of guests
Kwon H. et al. [26]	air Lines	Extracting features that affect customer satisfaction and dissatisfaction
Chatterjee S. et al. [8]	Health services/health products	Extracting effective features and added services and analyzing various feelings in the context of each in order to predict customer satisfaction
He L. et al. [19]	medicine industry	Automatic extraction of latent themes from online pharmaceutical reviews of customer dissatisfaction
Li N. et al. [27]	Products - smart phone	Extracting the required features of the products - sentiment analysis - investigating the relationship between the extracted features and customer satisfaction
Davras O. and Caber M. [10]	hotel management	Extracting effective features in services on customer satisfaction and dissatisfaction

2.3 Text mining and semantic network analysis

Text mining is defined to use of information retrieval and extraction using natural language processing techniques to discover unknown useful patterns and discover knowledge in the text [15, 14, 3, 21], in general, the process of text mining includes the stages of data collection, extraction and analysis and includes a management information system [15, 14, 3].

Finding a suitable model for text analysis is a complex process in text mining and information retrieval. Based on Kusumawati's model theory [25], models are representations of objects that depict selected characteristics and also depend on the pragmatic goals of the model creator. In this study, a rich framework for the analysis of semantic network structures is discussed to refine and adapt them for semantic analysis, which is mainly focused on qualitative aspects [12].

By modelling the text as a system of signs represented by different related words, we obtain a formal structure of word relations that corresponds to the model $G = (V, E)$ with $E \subseteq [V] \wedge 2$. According to graph theory [18], V is a set of nodes that represent different words as unique keys, and E is a set of edges that represent relationships between

different words to record a certain system of signs in a text. Therefore, G is a formal model of a generalized semantic network that consists of different words related to each other [12] with this model, the inherent meaning of the context can be identified. In addition, it is possible to analyze the impact of a specific word through the analysis of frequency and clusters of understanding words and how a specific word affects relationships between groups [3, 19] Semantic network analysis, as a method for Text analysis provides significant theoretical and methodological foundations for describing the semantic nature of texts in the field of tourism [3, 19, 38] Table 2 lists the studies that have adopted the semantic network analysis approach:

Table 2: Previous studies with the semantic network analysis approach

Authors	The area under investigation	Application of online comments in the article
Ban H. et al. [3]	hotel management	Investigating and extracting key features and evaluating the structural relationship of features
Atabay and Çizel [2]	hotel management	Analyzing opinions about 3 hotels - evaluating customer satisfaction by analyzing sentiments from texts - extracting features expected by customers
Oh M. and Kim S. [31]	food industry	Investigating dimensions in online comments to identify new quality indicators
Shadiyar A. et al. [32]	air Lines	Comparing three airlines by analyzing the relationship between customer experience and satisfaction
Zhang X. and Kim H. [39]	hotel management	Extracting the characteristics perceived by the customer and examining their relationship with customer satisfaction
Fu W. et al [15]	Restaurant	Extracting effective features in customer experience and investigating their impact on customer satisfaction
Fu W. et al. [14]	hotel management	Extracting effective features in customer experience and investigating their impact on customer satisfaction
Kim Y. and Kim H [22]	hotel management	Extracting customer preference features and examining their relationship with customer satisfaction
Wei S. and Kim H. [37]	hotel management	Extracting the desired features of customers and checking their relationship with customer satisfaction and dissatisfaction

3 Methodology

3.1 Data collection

The studied data was collected from the TripAdvisor website. This site has a large community of travellers around the world and has a huge amount of information that researchers can use to help tourism service providers and future travellers [20]. Considering that luxury hotels play a major role in attracting tourists, especially international tourists, in this study, online opinions of five-star hotels in Iran have been collected. A web scraping approach has been used to collect online comments using the libraries available in Python 3.9. The title and content along with the score that the user gave to the hotel for their satisfaction were collected. The title, along with the main text, was considered as content to be reviewed.
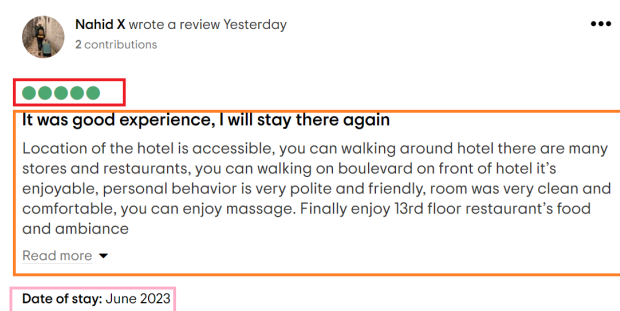

Figure 1: Components of an online comment

9907 comments from 30 five-star hotels in Iran that were collected from the site. Since English is known as an international language, to unify the studied language, the focus has been placed on the comments in English, and a total of 7417 online comments were extracted. All comments are collected until June 2023. This research includes two stages of data analysis; The first part includes text mining and exploratory analysis. The purpose of this type of analysis is to extract effective features in understanding customer experience or in other words customer preferences.

In the second part of the analysis, the relationship between the extracted features and their impact on customer satisfaction is modelled.

3.2 Text mining and exploratory analysis

In this section, the analysis is based on the research method in previous studies [15, 14, 3, 21, 33, 38]. The analysis was performed in three stages. The first step includes text preprocessing. Stop and repetitive words along with signs and symbols, pronouns, prepositions and meaningless words were removed from the texts and the generated data only includes meaningful words in the customers' experience. Word frequency was calculated for each word. According to the number of reviewed texts and by referring to the relevant articles, 100 words were extracted from the most repeated words in frequency analysis, and by checking the relationship of the words and the placement of the words in the same sentences, a co-occurrence matrix was created.

The co-occurrence matrix of words was given as input information to UCINET6.0 software to discuss the semantic network of keywords with Betweenness centrality and Eigenvector centrality to detect the power of nodes [14, 3] Special vector centrality of the concept of centrality It expands connectivity by not only considering the number of connected words, but also the importance of a connected relationship [19, 26, 32] Freeman degree centrality calculates the number of connections of a node, and when a word is most connected, will probably be central. The connection of the nodes shows the influential and dominant position of the nodes. After removing the less important words, different clusters are created by concord analysis. In fact, Concours analysis is a method for repeatedly analyzing correlations to search for certain levels of similar groups [17].

3.3 Results of exploratory analysis

3.3.1 Frequency analysis

Text mining is used for unstructured text data to structured data and the collected words were sorted according to frequency, thus 100 high-frequency words were extracted and organized for each of the groups, which are shown in Table 3. The collected keywords were classified as matrix data and used for semantic network analysis.

3.4 Semantic network analysis

Semantic network analysis is a social network analysis that emphasizes the relationship between actors when understanding the systemic structure of society [35]. Semantic network analysis shows how individual nodes are related to each other as a relational relationship and in the process, visually shows which nodes are used to construct a discourse [14]. In semantic network analysis, a "node" is a "word" that is primarily addressed in each study, not an individual. "Communicative relationship" refers to the "communicative relationship between words". The unique strength of semantic network analysis is its focus on interactions between words and their co-occurrence to produce meaning.

In this study, two parameters are used to analyze the connections and the role of keywords: Freeman rank center and eigenvector centrality. In Table 3, the calculated values of Freeman's eigenvector and centrality were calculated and among 100 words, 44 words with higher eigenvector values and centrality were selected. Then, according to the defined window, the co-occurrence matrix of the words in each comment was extracted. This co-occurrence matrix was used as the input of UCINET6.0. NetDraw was used to visualize the results. Then, the relationship and link between the words were identified by concord analysis. The greater the similarity of the connection model and the greater the degree of structural equivalence of words with each other, the greater the similarity of keywords in each cluster [37] This study uses node blocks based on the correlation coefficient of the matrix of co-occurring keywords to form clusters containing words similarly identifies [17].

Five clusters were extracted from the interactive analysis of Conkur. The naming of these groups was done with the help of past studies that were conducted in the field of the hotel industry based on the group of words extracted in each cluster: food service [21], service [14] and location, staff behaviour [21], Facilities [15] and purposes (features) of travel [3, 21] food service group includes "good", "restaurant", "food", "quality", "delicious" and "buffet" which indicates the quality of food and food service in the restaurant and buffet. Service and location include the keywords "Breakfast", "Service", "Nice", "Great", "Time", "Location", "Beautiful", "Perfect", "Café" and "Shiraz". The employee behaviour group consists of the words "staff", "excellent", "friendly", "helpful", "reception", "polite", "professional", and "kind", which is one of the effective features in customer satisfaction. It is in the hotel industry. The facilities group also includes "room", "clean", "comfortable", "lobby", "floor", "big", "bathroom", "bed", "modern", "well" and "pool", which the owner of the facilities that the hotel offers to travellers; And the last group which is

Table 3: The list of 100 repeated words along with the average degree and the value of the eigenvector

Word	Freeman's middle grade	Special vector degre	Word	Freeman's middle grade	Special vector degree
Good	2.670	0.437	First	1.739	0.036
Room	2.670	0.264	guest	1.513	0.010
Staff	2.670	0.328	standard	1.895	0.024
Breakfast	2.670	0.287	choice	1.348	0.033
Service	2.670	0.203	new	2.593	0.035
Nice	2.670	0.225	*cafe*	2.622	0.060
Great	2.670	0.189	dinner	2.316	0.039
Restaurant	2.670	0.178	free	2.117	0.036
Excellent	2.670	0.171	facility	2.019	0.038
Food	2.670	0.182	wonderful	1.644	0.032
Clean	2.670	0.228	international	2.536	0.028
Time	2.670	0.080	*Stay*	2.670	0.135
Friendly	2.670	0.178	*buffet*	2.319	0.061
View	2.371	0.152	small	1.835	0.027
Night	2.670	0.054	*Pool*	2.437	0.054
Location	2.670	0.130	family	1.486	0.025
Tehran	2.670	0.092	high	1.772	0.022
City	2.164	0.104	coffee	2.237	0.026
Place	2.670	0.089	Next	2.670	0.025
Hotel	0.000	0.00	close	1.770	0.031
Helpful	2.594	0.126	special	1.621	0.027
Bad	1.663	0.057	traditional	1.256	0.028
Beautiful	1.739	0.084	hour	0.735	0.007
Comfortable	1.739	0.101	lovely	1.444	0.029
Day	1.855	0.030	Taxi	1.666	0.018
Lobby	2.618	0.076	luxury	1.686	0.025
Floor	2.363	0.064	expensive	1.620	0.025
Reception	2.670	0.061	center	2.138	0.029
Big	1.739	0.063	problem	1.661	0.016
Quality	2.580	0.064	*shiraz*	2.591	0.060
Perfect	2.371	0.055	Visit	1.315	0.025
Polite	1.663	0.073	Star	0.000	0.00
Experience	1.931	0.043	minute	0.370	0.003
Bathroom	1.589	0.054	team	0.837	0.019
Bed	2.412	0.038	Top	1.749	0.025
Old	1.104	0.036	manager	1.824	0.015
Trip	2.190	0.032	little	1.681	0.023
Modern	2.593	0.043	different	1.319	0.020
Professional	2.545	0.050	Way	1.869	0.016
Price	2.226	0.033	morning	2.126	0.020
Airport	2.294	0.038	thing	0.000	0.00
Business	2.578	0.036	mountain	0.324	0.014
Iranian	1.258	0.036	quiet	1.710	0.025
Area	2.630	0.034	available	1.315	0.019
People	1.663	0.034	Last	1.790	0.016
Large	1.841	0.044	fantastic	1.665	0.024
Well	2.670	0.091	hospitality	1.010	0.019
Delicious	2.565	0.070	variety	0.640	0.030
Spacious	1.401	0.057	internet	1.169	0.024
Kind	1.663	0.064	enough	2.648	0.028

named to travel goals includes "view", "night", "Tehran", "city", "trip", "price", "airport", "business" and "Area", which are related to travel.

3.5 Factor Analysis

Exploratory factor analysis has been used for a more detailed investigation and according to the number of keywords extracted to discover hidden variables. Factor analysis is used to discover the commonalities between keywords, and by using the variance between these keywords, it shows the relationship of variables in online comments. The purpose of this analysis is to reduce a large number of variables to smaller factors using the diagonal rotation process [3].

By semantic network analysis, 44 effective words were selected and given as input to factor analysis. A total of 20
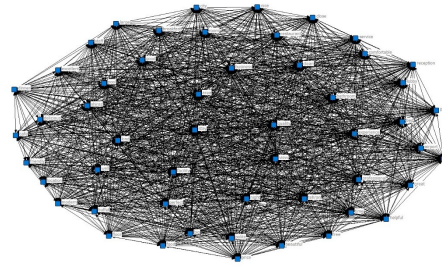
Figure 2: Graph of relationships between effective words - drawn with the help of word co-occurrence matrix
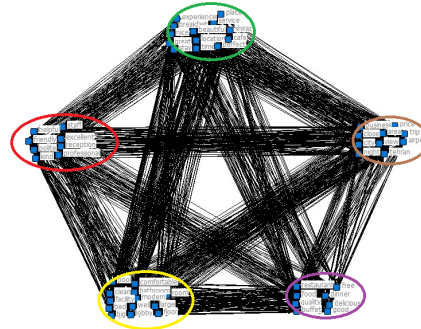

Figure 3: Visualization of word clustering using Kankur model

words in seven factors were extracted by repeating the factor analysis and by alternately removing the variables that correlated 0.4 in two groups or the univariate factors. In addition, the sampling adequacy measurement method and Bartlett's test of sphericity were used to confirm the appropriateness of the factor analysis method. The amount of KMO index was calculated equal to 0.704, which is acceptable considering that it is greater than 0.6 to explain the appropriateness of factor analysis [15, 14, 33]. Bartlett's test, the value of sphericity ($X2$) was 8244.843, ($p < 0.001$). In addition, factors should also have eigenvalues greater than 1.0. The percentage of total variance explained by the same factors is 42.682.

Based on the keywords in each factor, factors with titles of room facilities (RF), staff behaviour (SB), breakfast (B), travel purpose (TP), restaurant (R), lobby service (LS), city beauty (C) It has been tagged. The nomenclature is based on the common features and juxtaposition of the words. Factor one includes "bed", "bathroom", "room" and "comfortable", which all indicate the facilities of the room. The second factor includes "staff", "helpful" and "friendly", which are behavioural characteristics of employees. The third factor consists of "breakfast", "free" and "buffet", which was named breakfast because of the dominant method of serving breakfast in a buffet. The fourth factor includes "Tehran", "business" and "trip", which represents the purpose of the trip. The fifth factor, includes "food" and "restaurant", which was generally assigned the name of the restaurant. The sixth factor is the words "service", "lobby" and "café", which was labelled with the name of lobby services. The last factor was named after the words "beautiful" and "city".

3.6 Linear regression

Regression analysis was performed to determine how the independent variables influence the dependent variable. This analysis consists of seven independent variables, which are: room facilities (RF), staff behaviour (SB), breakfast (B), travel purpose (TP), restaurant (R), lobby services (LS), city beauty (C)) and an independent variable which is customer satisfaction and the score given by the customer based on his perceived experience to hotel services. As can be seen in Table 5, the overall explained variance ($R \wedge 2 = 0.06$) is equal to six, which indicates a relatively low correlation between independent and relatively dependent variables. In several articles that have been extracted with the approach of text mining, independent variables have encountered this problem [15, 14, 3, 21, 33, 38].

In these studies, such low explanatory power was considered possible due to the nature of exploratory research and the fact that it is impossible to include all related variables in the model.

The results of regression analysis show that "Room's service ($\beta = -0.468$, $p < 0.005$)", "Staff Behavior ($\beta = 0.225$, $p < 0.005$)" and "City ($\beta = 0.217$, $p < 0.005$)" The important predictors of customer satisfaction have

Table 4: Factor analysis results

Factors	Words	Loading	Eigen Value	Variance
Room Facility (RF)	Room	.595	2.429	7.945
	comfortable	.402		
	bathroom	.619		
	bed	.690		
Staff Behavior (SB)	Staff	.723	1.495	15.113
	Friendly	.663		
	Helpful	.636		
Breakfast (B)	Breakfast	.669	1.325	21.156
	free	.401		
	buffet	.736		
Trip's Purpose (TP)	Tehran	.562	1.266	26.833
	business	.737		
	trip	.598		
Restaurant (R)	restaurant	.711	1.120	32.345
	food	.699		
Lobby Service (LS)	service	.445	1.107	37.551
	Lobby	.544		
	Café	.731		
City (C)	city	.630	1.075	42.682
	beautiful	.627		
KMO (Kaiser–Meyer–Olkin) = 0.704; Total variance (%) = 42.682; Bartlett chi-squared (p) = 8244.843 ($p < 0.001$)				

Table 5: Results of linear regression model - relationship between factors and customer satisfaction

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	Beta	Std. Error	Beta		
Constant	4.266		.021	199.05	.000
Room Facility	-.468	-.212	.026	-17.700	.000
Staff Behavior	.225	.105	.025	8.969	.000
Breakfast	-.066	-.029	.027	-2.458	.014
Trip's purpose	.058	.023	.029	1.984	.047
Restaurant	.074	.032	.026	2.789	.005
Café	-.049	-.017	.033	-1.468	.142
City	.217	.083	.030	7.196	.000
a. Dependent Variable: overall satisfaction rating (Customer Satisfaction-CS); $R^2 = 0.06$; Adjusted R^2=0.059; F=65.936					

been; The facilities of the rooms have a major and negative role with the level of customer satisfaction, which indicates that dissatisfaction in this matter has been raised in online comments. After that, the factors of Restaurant ($\beta = 0.074$, $p < 0.05$), Breakfast ($\beta = -0.066$, $p < 0.05$) and Trip's purpose ($\beta = 0.058$, $p < 0.05$) play a role in the predictability of customer satisfaction. have had, and also regarding breakfast, the services provided have caused low customer satisfaction. Café factor ($\beta = 0.049$, $p > 0.05$) did not have an effective role in predicting customer satisfaction. The linear regression equation is given below:

$$CS = -0.468\,(RS) + 0.225\,(SB) + 0.217\,(C) + 0.071\,(R) - 0.066\,(BB) + 0.047(TP)$$

4 Conclusion

4.1 Discuss

In the age of digitization, collecting and analyzing customer feedback can ensure professionals gain timely insights from their customers and then provide a quick response to meet changing customer demands. This study was conducted to evaluate the experience and satisfaction of customers of five-star hotels in Iran based on online reviews. Therefore, this study has presented a new research design in a combination of qualitative and quantitative analysis of online comments to investigate the preferences affecting customer satisfaction in Iranian hotels.

To carry out this study, several steps have been carried out. First, keywords were extracted by text mining and the frequency of words in online comments was calculated, and 100 repeated words were selected. By analyzing the semantic network, the special vector and the middle vector of effective keywords were analyzed. 44 top words were selected and used as input for concours analysis to extract similar groups. These features extracted from online reviews

are a prerequisite for improving service quality and increasing customer satisfaction to gain a competitive advantage. Conkur's interactive analysis clustered the words into five groups. These five groups were named "food services", "location and services", "employee behaviour", "facilities" and "travel purpose".

Factor analysis was used to identify the hidden factors more precisely. In 20 words, they were categorized into seven factors named "room facilities", "staff behaviour", "breakfast", "travel purpose" "restaurant", "lobby services" and "beauty of the city". The obtained seven factors, in addition to the correlation with the output of Concor clustering, have also provided a more accurate classification. A linear regression model was implemented to investigate the effect of factors on customer satisfaction. The highest beta was shown to belong to room facilities such as room, bathroom, bed and room comfort, which hurt customer satisfaction. It indicates that this factor is the cause of customer dissatisfaction and the negative experience perceived by the customer. For example, the customer stated that "Rooms are not clean" or "Our room toilet door handle was broken". The second factor with a high beta that has a major positive effect on customer satisfaction is the behaviour of hotel staff, which is included in the intangible services provided by hotels. Ban et al [3] also showed in their study the major positive effect of this factor on customer satisfaction in the hotel industry. After the behaviour of the employees, in this study, the third effective factor in the predictability of the customer satisfaction of City Beauty was found, which has a major positive effect on customer satisfaction. All the above three factors had a value of $p < 0.005$, which indicates the major effect of these three factors. The next two influencing factors are related to food services, one of which is classified as a restaurant, and the other is divided into breakfast according to the main services of hotels. Therefore, the fourth factor is the restaurant and the food provided in the restaurant, which has a positive effect on customer satisfaction. The fifth factor of breakfast, which is still in the food service category, despite the restaurant, has hurt customer satisfaction, which in this case also indicates the unpleasant experience perceived by the customer. In this case, in the reviews, sentences like "just 15-20 minutes we can use of breakfast buffet, although breakfast was very poor and very limited" and "breakfast buffet was ok if we don't consider this hotel as a 5 star ". It is visible. The last factor that has a positive effect on customer satisfaction is the purpose of the trip. Among the considered characteristics, the cafe factor has not had an acceptable effect on the predictability of customer satisfaction. According to these findings, hotel companies can provide strategies based on their weaknesses and also based on customer preferences.

5 consequences

5.1 Theoretical implications

One of the main consequences of this research is the expansion of the use of semantic networks and factor analysis in the Iranian hotel industry with a text mining approach, which is related to the use of online surveys when studying customer satisfaction and understanding customer experience. Therefore, we claim that if hotels want to have a more accurate idea of the exclusive preferences of customers and their level of satisfaction, the analysis of online reviews is necessary.

One of the findings of this research is the selection of effective features in the hotel industry. The diversity of research in the field of hotel management has led to dealing with many simple and multi-dimensional features. In fact, analyzing online reviews has two advantages. Effective and optimal dimension reduction has been found on various features that have been discussed in various studies so far, and another point is to extract cases that may not have been discovered according to the study subject so far. These studies can be expanded in the field of one-dimensional variables and cover multi-dimensional variables.

According to our theoretical findings, the two factors of room facilities and staff behaviour had a high coefficient in customer satisfaction, where room facilities had a major negative effect and staff behaviour had a major positive effect. These numbers show that the features of the room such as bathroom, bed, and comfort play an effective role in the loss of customers because the hotels are five stars. On the other hand, the behaviour of employees has played a significant role in increasing customer satisfaction. In previous studies, some researchers also pointed out the importance of these aspects to improve the attractiveness of the hotel. One of the new findings of this research is the extraction of an effective factor called the beauty of the city, which has a positive effect. has had on the satisfaction of travelers and it shows that the pleasure that people get from seeing the beauty of the city extends to their perceived experience of hotel services. Of course, in past research [4], the importance of the subject under the name of location on customer satisfaction has been investigated, but here the effective factor of location on the satisfaction of city beauty customers was identified. Another finding that this research brought is that in the research of meals, they are placed in one group [3] if the results obtained from this study show that due to the influence to some extent, These things are the same, but their types are different. Paying attention to these issues makes it possible to distinguish the customer's preferences more accurately. Another final finding shows that as it was said based on the definition of

satisfaction theory, if the level of service is higher than the level of expectations, it leads to customer satisfaction, it can be concluded that people on business trips focus on goals They have a lower level of expectations from the services and this increases the satisfaction of the customers.

5.2 Managerial implications

The most important managerial consequence of this research is the discovery of the major negative impact of room facilities. Managers should pay more attention to the fact that due to five-star hotels, customers expect rooms with higher standards. In addition, the modernization of these spaces and the use of smart facilities in the rooms can increase customer satisfaction. The second consequence is that due to the impact of employee behavior on customer satisfaction, it is possible to make customers happy by continuously training employees in the direction of customer orientation and increasing their capabilities, such as hospitality etiquette and also by teaching living languages of the world. . The third consequence is a recommendation that can be given to the managers for hotel location. It means choosing the city and the surroundings of the hotel. that both the chosen city has enough beauty to attract tourists and beautify the area around the hotel. The fourth recommendation that the managers should consider in their decisions at the level of five-star hotels is the breakfast provided in the buffet, including free and detailed presentation, which can not only solve the dissatisfaction of the customers, but also can increase customer satisfaction; And the last point is attracting customers based on the purpose of their trip, who can offer special offers to customers based on their trip.

5.3 Limitations and future research

Considering that this research has created valuable insight into the weaknesses of the hotel industry by analyzing online reviews, but it still has some limitations. The first limitation is that because these hotels are located in Iran, in addition to reviewing comments in English, Farsi and all other languages should be considered for analysis. The second issue is the limitation of the studied data. In this research, only the data of five-star hotels from TripAdvisor has been examined, and it is necessary to increase the number of comments by increasing the number of hotels and other data sources to generalize the cases. Thirdly, the data has been examined in general, but in the future, it is better to examine the satisfaction and dissatisfaction of customers separately according to each hotel to obtain a better understanding of the customer experience and stronger strategies to increase profits. He presented the profitability and profitability of hotels.

In future research, due to the low overall variance in the regression model, a wider effort should be made in this area because the comment texts contain a variety of words that are actually associated with a meaning and a level of service, hence for the above Taking this number and covering more of the existing variables, the set of similar and synonymous words can be used to reduce the diversity of words in extracting key words and factor analysis. Another thing that is recommended in future research is to examine the country where people live, which plays a role in defining people's preferences, and it is necessary to personalize the interpretation of the results by taking into account the place where people live.

References

[1] A. Ahani, M. Nilashi, E. Yadegaridehkordi, L. Sanzogni, A.R. Tarik, K. Knox, S. Samad, and O. Ibrahim, *Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels*, J. Retail. Consumer Serv. **51** (2019), 331–343.

[2] L. Atabay and B. Çizel, *Comparative content analysis of hotel reviews by mass tourism destination*, J. Tourism Serv. **11** (2020), no. 21, 147–166.

[3] H.J. Ban, H. Choi, E, K. Choi, S. Lee, and H.S. Kim, *Investigating key attributes in experience and satisfaction of hotel customer using online review data*, Sustainability **11** (2019), no. 23, 6570.

[4] J. W. Bi, Y. Liu, Z. P. Fan, and J. Zhang, *Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry*, Tourism Manag. **77** (2020), 104006.

[5] S. Borhanian and V. Reza Mirabi, *Factors affecting customer satisfaction in the private sector hotels in Qom case study: Parsia Grand Hotel (4 starsT)*, UCT J. Manag. Account. Stud. **5** (2017), no. 1, 25–33.

[6] K. Celuch, *Customers' experience of purchasing event tickets: mining online reviews based on topic modeling and sentiment analysis*, Int. J. Event Festival Manag. **12** (2021), no.1, 36–50.

[7] Y.C. Chang, C.H. Ku, and D.D. Nguyen, *Predicting aspect-based sentiment using deep learning and information visualization: The impact of COVID-19 on the airline industry*, Inf. Manag. **59** (2022), no. 2, 103587.

[8] S. Chatterjee, D. Goyal, A. Prakash, and J. Sharma, *Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application*, J. Bus. Res. **131** (2021), 815–825.

[9] S. Chatterjee and P. Mandal, *Traveler preferences from online reviews: Role of travel goals, class and culture*, Tourism Manag. **80** (2020), 104108.

[10] Ö. Davras and M. Caber, *Analysis of hotel services by their symmetric and asymmetric effects on overall customer satisfaction: A comparison of market segments*, Int. J. Hospital. Manag. **81** (2019), 83–93.

[11] K. Ding, W.C. Choo, K.Y. Ng, S.I. Ng, and P. Song, *Exploring sources of satisfaction and dissatisfaction in Airbnb accommodation using unsupervised and supervised topic modeling*, Front. Psycho. **12** (2021), 659481.

[12] P. Drieger, *Semantic network analysis as a method for visual text analytics*, Proc.-Soc. Behav. Sci. **79** (2013), 4–17.

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

[14] W. Fu, E. K. Choi, and H.S. Kim, *Text mining with network analysis of online reviews and consumers' satisfaction: A case study in Busan Wine Bars*, Information **13** (2022), no. 3, 127.

[15] W. Fu, S. Wei, J. Wang, and H.S. Kim, *Understanding the customer experience and satisfaction of casino hotels in Busan through online user-generated content*, Sustainability **14** (2022), no. 10, 5846.

[16] A. Ghorbani, H. Mousazadeh, F. Akbarzadeh Almani, M. Lajevardi, M.R. Hamidizadeh, M. Orouei, K. Zhu, and L.D. Dávid, *Reconceptualizing customer perceived value in hotel management in turbulent times: A case study of Isfahan Metropolis five-star hotels during the COVID-19 pandemic*, Sustainability **15** (2023), no. 8, 7022.

[17] M. Guha Majumder, S. Dutta Gupta, and J. Paul, *Perceived usefulness of online customer reviews: A review mining approach using machine learning & amp; exploratory data analysis*, J. Bus. Res. **150** (2022), 147–164.

[18] F. Harary, *Graph Theory*, CRC Press, 1969.

[19] L. He, D. Han, X. Zhou, and Z. Qu, *The voice of drug consumers: Online textual review analysis using structural topic model*, Int. J. Envir. Res. Public Health **17** (2020), no. 10.

[20] H.J. Ban, *Semantic network analysis of hotel package through the big data*, Culinary Sci. Hospital. Res. **25** (2019), no. 2, 111–119.

[21] R. Khorsand, M. Rafiee, and V. Kayvanfar, *Insights into trip advisor's online reviews: The case of Tehran's hotels*, Tourism Manag. Perspect. **34** (2020), 100673.

[22] Y.J. Kim and H.S. Kim, *The impact of hotel customer experience on customer satisfaction through online reviews*, Sustainability **14** (2022), no. 2.

[23] F. Kitsios, M. Kamariotou, P. Karanikolas, and E. Grigoroudis, *Digital marketing platforms and customer satisfaction: Identifying ewom using big data and text mining*, Appl. Sci. **11** (2021), no. 17.

[24] S. Kuhzady and V. Ghasemi, *Factors influencing customers' satisfaction and dissatisfaction with hotels: A text-mining approach*, Tourism Anal. **24** (2019), no. 1, 69–79.

[25] A. Kusumawati, F. Aprilia, and Y. Abdillah, *Analyzing influence of electronic word of mouth (EWOM) towards visit intention with destination image as mediating variable: A study on domestic visitors of museum angkut in batu, Indonesia*, Eurasia: Econ. Bus. **19** (2019), no. 1, 50–57.

[26] H.J. Kwon, H.J. Ban, J.K. Jun, and H.S. Kim, *Topic modeling and sentiment analysis of online review for airlines*, Information **12** (2021), no. 2, 1–14.

[27] N. Li, X. Jin, and Y. Li, *Identification of key customer requirements based on online reviews*, J. Intell. Fuzzy Syst. **39** (2020), no. 3, 3957–3970.

[28] E.E.V. Martínez and A.D. Cruz, *Big data in hotel companies: A systematic literature review*, Digital Transformation of the Hotel Industry , 2021, pp. 111–133.

[29] N. Naderi, *Proposing a model for data driven marketing in the smart tourism with a meta-synthesis approach*, Tourism Manag. Study **18** (2023), no. 61, 169–206.

[30] A. Nazarian, P. Atkinson, P. Foroudi, and D. Edirisinghe, *Factors affecting organizational effectiveness in independent hotels-the case of Iran*, J. Hospital. Tourism Manag. **46** (2021), 293–303.

[31] M. Oh and S. Kim, *Dimensionality of ethnic food fine dining experience: An application of semantic network analysis*, Tourism Manag. Perspect. **35** (2020), 100719.

[32] A. Shadiyar, H.J. Ban, and H.S. Kim, *Extracting key drivers of air passenger's experience and satisfaction through online review analysis*, Sustainability **12** (2020), no. 21, 9188.

[33] Y. Shen, J. Zhou, A.A. Pantelous, Y. Liu, and Z. Zhang, *A voice of the customer real-time strategy: An integrated quality function deployment approach*, Comput. Ind. Engin. **169** (2022), 108233.

[34] S. Tao and H.S. Kim, *Online customer reviews: insights from the coffee shops industry and the moderating effect of business types*, Tourism Rev. **77** (2022), no. 5, 1349–1364.

[35] M. Varmaghani, *Designing Brand Intelligence Model in Tourism Startups: A Mixed Approach*, Tourism Managment Study, 2023.

[36] W. Wang and R. Rada, *Structured hypertext with domain semantics*, ACM Trans. Inf. Syst. **16** (1998), no. 4, 372–412.

[37] S. Wei and H.S. Kim, *Online customer reviews and satisfaction with an upscale hotel: A case study of Atlantis*, Palm Dubai Inf. **13** (2022), no. 3, 150.

[38] A. Yucel, A. Dag, A. Oztekin, and M. Carpenter, *A novel text analytic methodology for classification of product and service reviews*, J. Bus. Res. **151** (2022), 287–297.

[39] X. Zhang and H.S. Kim, *Customer experience and satisfaction of disneyland hotel through big data analysis of online customer reviews*, Sustainability **13** (2021), no. 22, 12699.