A New Approach in Marketing Research: Identifying the Customer Expected Value through Machine Learning and Big Data Analysis in the Tourism Industry

Ali Ghasemian Sahebi, Rahil Kordheydari, Mohammad Aghaei*

Department of Management, Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran.

Correspondance Email: Alighasemiansahebi@modares.ac.ir; K.rahil@modares.ac.ir; M.aghaei@modares.ac.ir.*

Abstract

With the advent of the digital age and its entry into the business environment, marketing trends have undergone several changes, such as the expansion of data-driven marketing and the digitalization of the business environment that requires further investigation. As a result, the aim of this study is to present a novel approach in marketing research and identify value components among the large volume of customer feedback in virtual networks using machine learning, big data analysis, and a predictive marketing strategy. This research is applied in terms of purpose, qualitative-quantitative (mixed) in terms of method and descriptive-survey in terms of data collection that has used an inductive approach. For this purpose, the Iranian tourism industry and tourist areas of Tehran province was selected as a case study and 9325 comments from customers about hotels used in their travels to Tehran were collected from the Internet between the summer of 2020 to winter 2021 and using "data clustering" and "Association Rules Extraction" methods, the value components were extracted and RapidMiner software and Python programming language were used to perform data mining, text mining and big data analysis processes. In summary, the findings demonstrate that by employing big data analysis and machine learning, the process of "marketing research" can be performed with greater speed, accuracy, and extensiveness, as well as at a lower relative cost. The findings and suggestions of this study are used to inform marketing researchers, as well as to raise the awareness of managers, especially in hotel industry.

Keywords: Marketing Research; Customer Expected Value; Big Data; Machine Learning; Tourism Industry

Introduction

Today the dynamic conditions of marketing have caused the organizations to use the new technologies because marketing is experiencing deep reformations and developments at the present time to the extent that most economists believe that a new era namely the digital age is coming forth now (Gupta et al., 2017). Today, companies face challenges such as making quick decisions to improve productivity in the face of the big data problem; Because most of the production systems suffer from a lack of analytical smart facilities, they are not apt to manage their huge amounts of data (Lee, Kim & Lee, 2017). The new age describes a world that empowers people to move between digital and offline reality to live and manage it using
communication technology (Xu, David & Kim, 2018).

In the digital age, every business should try to use technological facilities to change and improve his/her market activities. This reformation is attained through utilizing such technologies as artificial intelligence, the internet of things (IoT), blockchain technology, virtual reality, mass media, etc (Sanayei, 2017). The advent of technologies such as artificial intelligence and machine learning has made marketing solutions far more effective than ever before and the high volume of instantaneous information production forces marketers and market analysts to use new tools for analysis to make appropriate decisions (Hsu et al., 2021). One of the most important technologies used in this field is artificial intelligence and machine learning (Xu et al., 2018). Machine learning is one of the new technological approaches that has greatly affected the field of marketing, so that 48% of marketing companies in 2019 admitted that they have used this approach to develop their service and facilitate forecasting. Also, consumer behavior analysis has testified with great accuracy (Kumar, Venkatesh & Rahman, 2021). Therefore, many companies have used machine learning algorithms to highlight hidden knowledge in consumer data in order to create a better experience for their customers, as well as improve their operational factors, such as speed and cost and time savings (Huang et al., 2021). Also, mechanized and computer-based market building has enabled machine learning algorithms to interact directly with markets and enable complex instant buy and sell orders based on complex decisions and identify potential customers with high potential (Kumar, Venkatesh & Rahman, 2021). Therefore, in general, most organizations use the benefits of machine learning techniques because of the ability of this approach to analyze large data sets and provide understandable results for marketers to optimize their marketing strategies and predict consumer behavior. In practice, machine learning charts a roadmap for present and future marketing (Ahmad, Jafar & Aljoumaa, 2019).

In every market, the customers have many choices to choose from, but they experience just a few of them in their decision making and the purchase of the product (Halawani, Soh & Muthaiyah, 2019). In digital transformation, instead of creating a new product, we are talking about creating value. Therefore, instead of creating a product, you should think about creating new value. Value for the customer is the benefit he receives in return for the money paid. This benefit can be financial, temporal, psychological, and so on. But the issue that is very important in the discussion of value creation is identifying the values expected by the customer and presenting them (Yen, Teng & Tzeng, 2020).

Marketing research processes in organizations are important to identify the values that customers expect. Marketing research, with the entry of businesses into the age of digital, has undergone changes and has taken new approaches (Ahmad, Jafar & Aljoumaa, 2019). With the expansion of the population and a large amount of organizational data in the market research process, the use of approaches that make this process easier and faster has become very important (Heidari, Valipour & Bakhtyiari, 2021). Therefore, among the multitude of new technologies and approaches created, it is necessary to identify cases that are applicable in this area and lead to improved results and optimal use of resources. One of the most reformation in this field is to access the technology of artificial intelligence, virtual channels, and big data in the process of identifying and analyzing the customers and market, etc. It will open new windows in this field. Therefore, these technologies and the approaches should be examined to see if they were effective, then develop throughout the organization (Czinkota et al., 2021).

Today many countries are busy working in the tourism industry as one of the most economical section in development. The tourism industry improves the prospects of economic growth in Iran and increases gross domestic product. The related industries will also be improved, and their positive results will illuminate the situation (Aghaei, Sahebi & Kordheydari, 2021a). For this very reason tourism industry is studied in the present research. Also, tourism in Tehran, as the capital and largest province of Iran, with its high share in tourism has been studied. one of the
most important reasons for increasing competition in this industry is the change in customer expectations and value-oriented of their choice of destination and hotels which encouraged managers in this field to become aware of the needs of their customers and try to know more than before (Kim, Ariza-Montes & Han, 2021), but there are still many other challenges in this field and the pandemy of Covid 19 put tragic effects in this industry.

In the literature of marketing, many researchers have emphasized that the use of potential capabilities of information systems would be one of the critical and important factors in success in the tourism industry (Raguseo, Neirotti & Paolucci, 2017). As investment in the tourism industry needs a huge amount of finance, studying and analyzing the parameters which the customers have in their mind or share on the internet is reasonable and necessary. Moreover, social media, web, and digital channels are a great help for the customers to take part in discussions in this field, they can express their viewpoints online and finally create a market by electronic Word of Mouth Marketing (eWoM) (Leong et al., 2019). Research done by Chong et al. (2018) was focused mainly on the role of comments and viewpoints on the internet by the users. The results showed that most passengers copied the comments of the passengers who had come there before, they reserved the hotels only based on what their friends or others have talked about the quality of the hotel. Kim, Ariza-Montes and Han (2021) by examining the effect of creating the expected benefits of customers on their behavior in choosing a hotel, they found that this effect is positive and significant and can shape the choice of tourists. Hair Jr and Sarstedt (2021) also found in a study that the advent of digital data and the methods used to analyze it revolutionized marketing research, and the vast amount of data provided countless opportunities to better predict and potentially explain consumer behavior to marketing researchers. In this regard, in general, it can be said that past research in the field of marketing research and customers (For example, Dwivedi et al., 2021) explored the future of digital networking and social networking; Kumar and Reinartz (2016) explored how to create value for customers; As well as research by Kumar, Venkatesh and Rahman (2021); Dolnicar and Ring, 2014; Suarniki and Lukiyanto, 2020; Saleem et al., 2021; Jiang and Hong, 2021) has been done more with a traditional approach, and research in the field of new technologies and approaches to the use of big data and machine learning in the new age (For example, Hsu et al., 2021) who segmented customers using machine learning online; Kumar, Venkatesh and Rahman (2021) who used machine learning and big data in retail; As well as research by Grimmer, 2015; Zhou et al., 2017; Beam and Kohane, 2018; Xie, Qian and Wang, 2021), each deal with a specific topic in these areas and research that connects the two areas is very limited, and therefore many research gaps in this the subject is felt.

The study of previous researchers showed that due to the novelty of the use of machine learning in marketing, there will be many opportunities and threats in the field of marketing research that is influenced by the consequences of the digital age, and research in this area is just the beginning, and there is a wide field for research in this field and there are many study gaps. This situation is very obscured in Iran to the extent that independent research especially in marketing via the approach of artificial intelligence in big data has not been done yet. There are many potentialities in this field in our country that will make a large amount of income for the country. As these approaches are new, they need to be examined for their validity and positive effects. Because the values of the customers are relative and the difference of values exist in distinct groups, the analysis of values expected by customers is necessary and must be done separately with regard to the near and close relation of the tourism industry and hoteling. The expectations of customers are important, and they should be always measured carefully. The pandemic of covid-19 and its great effects in this industry was also very important as well. The expectations of customers will be the most effective parameter to measure to discover the positive or negative experiences of customers from their traveling. Analyzing their viewpoints concerning future traveling is also of great importance. Therefore, according to the gaps expressed in this area, the purpose of this study is to describe a new approach in marketing research and identify value components from the
opportunities but also place more emphasis on orientation in order to achieve sustainable advantage.

It should be emphasized that marketing research studies are gradually becoming more and more complex. In addition, the impact of unstable and difficult-to-measure factors on the behavior of market players, and changes in research conditions, including the creation of artificial and natural health for experiments, exacerbate the problems of interpreting research results. Also, the development of new technology applications is transforming marketing research by changing from classical analysis to new technology-based study (Wang et al., 2021). Therefore, as we enter the new era, due to increasing competition, market dynamics, changing customer tastes, and the emergence of large amounts of digital data as big data, more and more marketing research is important for companies and the importance of adopting marketing research to new market needs has increased. To use today's big data, marketing science must use disciplines such as data science, machine learning, word processing, audio processing, and video processing (Miklosik & Evans, 2020).

In this regard, Hair Jr and Sarstedt (2021) in a study, showed that using machine learning in data analysis and measurement can create new opportunities for marketing, such as speed and accuracy, but may lead to increased costs. Soltani Zenozi (2019) in another study has examined the role of big data in marketing and business. Research findings show that the use of big data improves the marketing and sales process by optimizing and personalizing marketing strategies and increases customer satisfaction and loyalty, thus balancing an organization's products with customer needs for both business and leisure. It will bring more profit and value to the customer.

Customer Expected Value

In today's increasingly competitive world, successful companies will be more likely to satisfy their customers (Mansouri Moayyed, kordheydari & Ghasemian Sahebi, 2020). Companies that are not just looking for short-term sales and long-term customer satisfaction by providing goods and services with superior and...
distinctive value of their efforts (Aghaei, Sahebi & Kordheydari, 2020). In this highly dynamic market, the customer expects the organization to offer the most value at the most appropriate price, and organizations are constantly looking for new ways to innovate in creating and delivering value, and even customer value as a "future source of competitive advantage" are named (Jiang & Hong, 2021).

Regardless of the type of industry and product, organizations strive to gain a privileged position in the mind of the customer, create lasting relationships with him, increase loyalty and satisfaction and optimally meet the needs and wants of the consumer market, to produce and supply goods and services tailored to the needs Market and offer price for their product in exchange for material, human and information resources (Aghaei, Sahebi & Kordheydari, 2021a). Customers also get the desired product by paying these amounts to meet their needs. But the relationship created to sell a product on the one hand and buy it on the other is not only influenced by the financial factor of the exchange (Suarniki & Lukiyanto, 2020). The consumer of goods or the recipient of services connects to the organization and the product through a comprehensive and complex process that goes far beyond just purchasing the product, and his evaluation of this process goes beyond the price paid and the functional characteristics of the product (Dolnicar & Ring, 2014). In other words, the value that the buyer attributes to this process includes several factors along with the price and function of the product, which is referred to as "value from the customer's point of view" and as one of the key factors in achieving sustainable competitive advantage in today's market-oriented and customer-oriented world has been considered by researchers and managers of organizations (Yen, Teng & Tzeng, 2020). The emergence of new tools for communication with customers, including social networks, has created new challenges and opportunities in terms of how to introduce and deliver value to the organization's stakeholders (Ghasemian Sahebi, Moshabaki & Khodadad Hosseini, 2018). The concept of value has been used in different contexts with different definitions. From the customer's point of view, value is what the customer receives in return for the price he pays (Liu, 2021). Perceived value has been described as a major pioneer of consumer behavior based on the value-attitude-behavior approach in the tourism industry (Kim, Ariza-Montes & Han, 2021). In fact, the brand of tourism destination is tourists' perception of their travel experience about the value of the brand and the emotional value they place on the brand (Aghaei, Sahebi & Kordheydari, 2021b), and today these brandings are considered as a company's assets and affect the company's income (Kumar and Reinartz, 2016). Therefore, considering the importance of customer expected value in the field of tourism, in this study we have studied this issue.

Based on researches, it can be said that due to the Customer Value Proposition (CVP), customers enjoy the company's services and also create value for it by partnering with the company (An & Han, 2020). Tukiran, Tan and Sunaryo (2021) in a study found that quality, expectations and value are perceived by the customer and affect customer satisfaction.

**Big data**

Today, organizations and businesses collect and store vast amounts of data; They are hopeful to use them in the future to support managers' decision-making process (Amado et al., 2018). The term "big data" refers to a set of data that is at a high level in terms of speed, volume, and variety and is not processed with traditional techniques and tools. Today, due to the expansion of online space and increasing the volume of customer data, big data has become especially important for brands to analyze customer behavior (Kumar, Venkatesh & Rahman, 2021). Big data can change the landscape of innovation by effectively and efficiently increasing the fit between consumer preferences and product features (Ghasemaghaei & Calic, 2020). Nowadays, big data is everywhere; Whether in the form of structured data, such as traditional organizational databases (e.g., customer relationship management system) or unstructured data, which are new communication technologies and platforms that the user can develop or edit (e.g., texts, images, and videos) (Lansley & Longley, 2016). Two important sources of data today are social media and mobile applications, which studies show have a great impact on customer
Meaningful patterns from documentaries (Ghasemaghaei & Calic, 2020). Text mining of mass media pages helps us to get a better understanding of the customer. In marketing, text mining helps us to detect future sales, manage customer communication, getting the knowledge of our competitors, improve the brand establishment, the analysis of customer emotions, and making great decisions based on a fundamental understanding (Xie, Qian & Wang, 2021). The purpose of this analysis is to predict the future, especially in the field of marketing (Kumar, Venkatesh & Rahman, 2021). Predictive analysis allows marketers to use the data that customers generate while interacting online with systems, and to determine purchasing processes and user behavior models in order to create mechanized systems and determine user profiles for specific target markets (Hsu et al., 2021). But to ensure the use of artificial intelligence in marketing, you must first consider how to use and also the benefits of using this method.

To highlight the important role and advantages of using Artificial intelligence and machine learning in improving the market processes, Krafft, Sajtos & Haenlein (2020) in a study found that marketing changes in the new era should not be ignored and today the use of traditional approaches cannot meet the needs of marketing and artificial intelligence and new technologies should be used to improve marketing performance for customer satisfaction. Amir Khani and Mottaghi (2018) tried to explain how machine learning uses big data in making models for the purpose of detecting. Their study showed that the presented approach would be able to detect the production with the least error and the highest carefulness. It could eliminate lots of the problems related to analyzing complicated data in management systems of production. Ma and Sun (2020) in their study on the use of artificial intelligence and machine learning in marketing, found that this technology is required to meet the novel needs of marketing and would increase the quality and the outcomes of marketing.

**Methodology**

This study aims to identify the components of value among the large volume of customer feedback in virtual networks in the tourism industry using machine learning. In terms of applied purpose, the present study is a
Descriptive survey in terms of data collection method and qualitative-quantitative (mixed) in terms of implementation with an inductive approach. Methods and algorithms for textual big data analysis through artificial intelligence (machine learning) have been used to analyze the content of expected values of customers to, first, the customer expectations are determined and the application of the machine learning method in this process is examined, and second, the advantages that this method can bring over the traditional methods are evaluated. The study of this research is the tourism industry of Iran and tourist areas of Tehran province, and the data used in this research are the opinions of customers registered on the Internet, including sites and social networks, and their use of the hotel in their travels to Tehran. The dataset, due to the decline in tourism during the outbreak of Covid-19 virus, contains 9325 comments collected from summer 2020 to winter 2021.

RapidMiner software and Python programming language have been used to perform data mining, text mining, and big data analysis processes. The implementation stages of the research include three main stages that are described below.

Data Pre-Processing

Input data in text mining is often unstructured text documents; For this reason, input data must first be pre-processed. In this process, the text is first divided into words, symbols, or other meaningful elements called "Tokens" (Weiss, Indurkhya & Zhang, 2015). To extract the words of a text, the data is first converted to a line of words by removing punctuation and other non-text characters. Figure 1 "Word Cloud" shows tourists' comments about the hotel after data preprocessing.

Data clustering

In the next step, the "clustering" method is used to extract the "value" components from the large volume of customer feedback. Clustering is the process by which data is divided into groups whose members are similar to each other. A cluster is a collection of similar objects that are dissimilar to objects in other clusters (Raschka & Mirjalili, 2017). Figure 2 shows the data clustering process.

In text mining, clustering is a technique used to group similar documents (in one subject). Clustering texts is needed when the volume of textual documents is very large and the possibility of separating them based on the topic that each group of documents refers to is impossible or very difficult and time consuming for humans. One of the most popular methods is data clustering using the "k-means" method, which considers the similarity of textual documents and the relationship between keywords (Weiss, Indurkhya & Zhang, 2015). In this step, RapidMiner software is used. Figure 3 shows data clustering by RapidMiner software in this study.

Figure 1: Word Cloud of Comment Words in Text Data
In this study, to determine the number of clusters and clustering validation, the "Dunn index" method has been used due to better performance than other methods in big data (Kassambara, 2017). The "Dunn" index is presented with the aim of determining the amount of compactness (distance of members within the cluster with each other) and separation (distance of clusters from each other) and calculates the amount of compactness and separation with two criteria of "cluster distance" and "diameter". This criterion is defined by Equation 1, in which \( d(C_i, C_j) \) is the distance between clusters \( i \) and \( j \) and \( \text{diam}(C_l) \) represents the diameter of cluster \( l \) (Faezi Rad & Pouya, 2016).

**Equation 1.**

\[
DI = \min_{i+1 < j \leq K} \left\{ \min_{i+1 < j' \leq K} \left\{ \frac{d(C_i, C_j)}{\max_{1 \leq l \leq K} \text{diam}(C_l)} \right\} \right\}
\]
**Association Rules**

Association rules are an approach to data mining that is used to discover repetitive patterns, correlations, relationships, and causal structures in a variety of databases and to analyze and predict user behavior. By discovering the connections between data sets, these rules show the conditions that occur frequently in a data set and explain the presence of some items (entities) based on other items (entities) (Shalini & Lal, 2016). The purpose of association rules in text mining is to discover the relationships between words and to identify the rules that require the presence of a set of words in the text and the existence of other words.

In a database, if \( I = \{I_1, I_2, ..., I_m\} \) is a set of items, \( D \) is a set of transaction databases, each of which has an identifier, as well as a set of items such as \( T \subseteq I \).

If \( A \) is a set of items, a transaction like \( T \) contains \( A \), if we have: \( A \subseteq T \). An association law is expressed as \( A \Rightarrow B \) in which \( A \subseteq I, B \subseteq I, A \cap B = \emptyset, B \neq \emptyset, A \neq \emptyset \). An association rule states in its general form that if event \( A \) occurs, then event \( B \) also occurs. Symbol \( A \Rightarrow B \) is used to represent an association rule, and \( A \) is called the introduction of the rule and \( B \) is the result of the rule. The source of the production of association rules are also times that have been repeated enough that they are called "collection of recurring items" (Han, Pei & Kamber, 2011). Each rule of \( A \Rightarrow B \) has an indicator called "Support" that indicates the repetition rate of the concurrency. The Support value of rule \( A \Rightarrow B \) is calculated using Equation 2. The numerator is equal to the number of concurrences \( A \) and \( B \), and the denominator refers to the total number of events.

**Equation 2.** \( \text{Support (A \Rightarrow B) = } \frac{N(A \& B)}{N(A)} \)

Once the rules have been generated with a minimum value of Support, the probability of the rules being correct can also be calculated. This indicator, called the "Reliability", indicates what percentage of the time that Event \( A \) occurred, Event \( B \) also occurred. The Reliability of rule of \( A \Rightarrow B \) is calculated using Equation 3 (Lee, Kim & Lee, 2017).

**Equation 3.** \( \text{Confidence (A \Rightarrow B) = } \frac{N(A \& B)}{N(A)} \)

---

![Figure 4: Extraction of Association Rules by Rapidminer](https://example.com/figure4.png)

Apriori and FP-Growth are two of the most important and practical methods of association rule extraction algorithms. In this study, after examining these two methods and comparing...
their efficiency and speed, the FP-Growth method was used to extract the association rules (Esmailipour, 2016). Figure 4 shows how RapidMiner performs the process of extracting association rules in this research.

**Results and Discussion:**

In this section, the findings of the text mining of customers' comments about the hotel in Iran's tourism destinations are presented in the form of two parts: "Clustering of comments" and "Extraction of association rules".

**Findings of the "clustering" process**

In this section, to extract the value components expected by customers, the collected comments are clustered using the k-means algorithm, and to identify the topic or feature mentioned in each cluster, the most frequently worded words in that cluster are identified. The naming of each cluster is done according to the logical relationship between the words in each category of topics that have the most repetition (probability) (Guo, Barnes, & Jia, 2017).

The results of clustering tourists' comments along with eight "words" with the most repetition (probability) in each cluster to identify the topic or feature desired by the customer in each cluster are shown in Table 1. By applying the clustering algorithm to the tourists' comments, 20 important clusters were identified. Examining each cluster and the words that have the most repetition in that cluster, the topic of discussion in that cluster is determined. Of the 20 topics or features identified in this collection, after sorting the clusters, the first 11 topics refer to the positive components of value and the next 9 topics refer to the negative components of value from the tourists' point of view in using the hotel. It should be noted that the components identified in Table 1 are not presented in order of priority and are only located in different clusters, which include two parts of positive and negative components.

### Table 1: Topics Extracted from the Clustering of Comments and Eight Words with a Higher Probability in Each Cluster

<table>
<thead>
<tr>
<th>Topics from Clustering Comments</th>
<th>Words” with the most repetition (probability) in each cluster”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1: Other comments about the hotel</td>
<td>Comments</td>
</tr>
<tr>
<td>Cluster 2: Cleanliness and general hygiene of the hotel</td>
<td>room</td>
</tr>
<tr>
<td>Cluster 3: Personnel behavior</td>
<td>Hotel</td>
</tr>
<tr>
<td>Cluster 4: Proximity to shopping and entertainment centers</td>
<td>Hotel</td>
</tr>
<tr>
<td>Cluster 5: Proper bathroom and toilet</td>
<td>Bathroom</td>
</tr>
<tr>
<td>Cluster 6: Room cleanliness (bedding, tablecloths, etc.)</td>
<td>room</td>
</tr>
<tr>
<td>Cluster 7: The beauty of the space and the interior layout</td>
<td>Hotel</td>
</tr>
<tr>
<td>Cluster 8: Regional facilities for recreation</td>
<td>Bar</td>
</tr>
</tbody>
</table>
Findings of the “Association Rules” process
The purpose of identifying association rules is to explore the relationships between words in a large set of texts (customer feedback). In this method, hidden knowledge is extracted from the analysis of words or phrases that customers have used together and abundantly in their comments. In Table 2, the rules are extracted, and two indicators of Support and Reliability are presented.

Table 2: Results of Extracting Association Rules and the Value of Two Indicators of Support and Reliability

<table>
<thead>
<tr>
<th>N</th>
<th>Introduction (Law)</th>
<th>Result</th>
<th>Confidence level</th>
<th>Backup amount</th>
<th>Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Health</td>
<td>Accommodation, hotel</td>
<td>0.9626</td>
<td>0.2317</td>
<td>(Health) ➔ Accommodation, hotel</td>
</tr>
<tr>
<td>2</td>
<td>Behavior</td>
<td>Accommodation, hotel</td>
<td>0.9650</td>
<td>0.3901</td>
<td>(Behavior) ➔ Accommodation, hotel</td>
</tr>
<tr>
<td>3</td>
<td>Service</td>
<td>Accommodation, hotel</td>
<td>0.9679</td>
<td>0.2221</td>
<td>(Service) ➔ Accommodation, hotel</td>
</tr>
<tr>
<td>4</td>
<td>room</td>
<td>Accommodation, hotel</td>
<td>0.9686</td>
<td>0.5174</td>
<td>(Room) ➔ Accommodation, hotel</td>
</tr>
<tr>
<td>5</td>
<td>Restaurant</td>
<td>Accommodation, hotel</td>
<td>0.9689</td>
<td>0.3008</td>
<td>(Restaurant) ➔ Accommodation, hotel</td>
</tr>
<tr>
<td>6</td>
<td>Breakfast</td>
<td>Accommodation, hotel</td>
<td>0.9734</td>
<td>0.4023</td>
<td>(Breakfast) ➔ Accommodation, hotel</td>
</tr>
<tr>
<td>7</td>
<td>Star</td>
<td>Accommodation, hotel</td>
<td>0.9847</td>
<td>0.2726</td>
<td>(Star) ➔ Accommodation, hotel</td>
</tr>
</tbody>
</table>
Association rules are defined as "if → then". In other words, by using this method, hidden relationships and dependencies in the big data are revealed and the discovery of important and useful rules facilitates the decision-making process of managers by providing appropriate information. For example, according to the findings in Table 2, if the hotel is hygienic, then tourists will choose to stay in that hotel.

Discussion:
Today, every business operates on the basis of a "business model". One of the most important components of a business model is "value proposition to the customer", which represents the product or service that the company is committed to delivering to customers (Sheehan & Bruni-Bossio, 2015). On the other hand, the most important component that can provide optimal value proposition to customers and create a competitive advantage for the organization, is to recognize the values expected by customers and strive to achieve it. In general, it can be said that recognizing and providing the right "value" to the customer is one of the most important challenges for the success and profitability of any business. This issue, with the introduction of the business environment into the digital age and the change of tastes and expectations of customers and the intensity of competition, has become more important and needs further investigation. With the growth of the tourism industry, the issue of the values expected by customers to choose a destination and a hotel, has become very important, which has faced challenges in this area. Also, the prevalence of Covid-19 has added to the evolution of customer expectations and necessitated research in this area. Past research that has linked marketing research to new technologies in the tourism industry is very limited. Tourism and hoteling industry in Iran, especially in Tehran province, despite its high potential, has low growth; therefore, due to the lack of research in this field and considering the new technologies and approaches that have entered the field of marketing to assess the value from the perspective of customers, this study was done with the aim of identifying the values expected by customers using new marketing research approaches (including machine learning and big data analysis) in the tourism industry of Iran and the tourist areas of Tehran.

The results of this study have been obtained through the two methods of machine learning (including clustering of comments and extraction of association rules), and big data analysis to explore features or value components commensurate with the expectations of customers (tourists), using the analysis of their online comments on the expected values of hotels in Iranian tourist destinations through RapidMiner software and Python programming language. The results and suggestions of this study is applicable in informing marketing researchers with how to use new approaches in this field to better understand customers and provide optimal services to them and also in increasing awareness of managers, especially in hotel industry about the expected benefits of customers and how to provide optimal services to gain competitive advantage.

The results obtained using the first method (data clustering) show 20 important components (subject or characteristics) in the perceived values of tourists, which include 11 positive components of value and 9 negative components of value. Positive features or components of perceived value of tourists are: 1) Opinions of others about the hotel, 2) Cleanliness and general hygiene of the hotel, 3) Good treatment of hotel staff, 4) Proximity of the hotel to shopping and entertainment centers, 5) Quality and variety of breakfast, 6) Transfer service and room delivery and 11) Additional facilities such as telephone, refrigerator, high speed internet and so on.

Also, the negative characteristics or components of the perceived value of tourists are: 1) lack of hygiene and not cleaning the room, etc., 2) power outages in the rooms, 3) inadequacy of the hotel star with its quality, 4) delay in transportation, 5) Existence of annoying insects in the room, 6) Low quality of personal hygiene items such as towels, shampoo, etc., 7) Intrusion of annoying noises from outside into the room, 8) High cost of
entertainment outside the hotel and 9) High and inappropriate price of the hotel.

The results obtained using the second method (association rules) include 7 important rules that have been obtained by extracting hidden knowledge from the relationships between the expressions used in tourists’ comments. The findings of this section show that 7 issues are more important for customers about staying in a hotel, which are: 1) general health of the hotel, 2) attitude of hotel staff, 3) how and quality of hotel service to tourists, 4) quality and conditions of rooms, 5) Quality and conditions of the hotel restaurant, 6) Quality and variety of breakfast and 7) Number of hotel stars. The results of extracting association rules provide the opportunity for service providers to be able to use the opinions of their customers and attract more satisfaction of customers.

Finally, the use of machine learning in the analysis of customer data showed that this method can analyze very quickly, which is a very important component in today's competitive world. Also, due to the lack of human intervention in the processes, the accuracy of the analysis has been very high. Extensibility is another advantage of this method, because it is possible to analyze big data from this method, if humans are not capable of this level of analysis. If this method is used extensively and in large volumes, the cost of analysis will be reduced, which is a very vital factor for many organizations.

The results of the study on the "importance of customer service quality" are consistent with the research of Sánchez-Franco, Navarro-García and Rondán-Cataluña (2019). The results of Chong et al.’s (2018) research show the importance of the opinions discussed on the Internet about the hotel and indicate that most travelers consider the opinions of previous travelers as a source of information for decision making. The results of the present study also confirm this and one of the positive elements obtained in the clustering section is the review of previous travelers’ comments about the hotel. Also, the results of this research are in some way in line with the results of (Chong et al., 2018; Hair Jr & Sarstedt, 2021; Amirkhani and Mottaqi, 2018; Kauffmann et al., 2020; Soltani Zanozi, 2019; Krafft, Sajtos & Haenlein, 2020; Ma & Sun, 2020; Tukiran, Tan & Sunaryo, 2021; Liu, 2021; Hsu et al., 2021; Kumar, Venkatesh & Rahman, 2021). Regarding the importance of using the opinions of previous tourists in cyberspace and using the potential capabilities of data analysis in this field, it should be said that the research of Raguseo, Neirotti and Paolucci (2017) and Sánchez-Franco, Navarro-García and Rondán-Cataluña (2019) are consistent with the approach of the present study. The research of Gil-Soto et al. (2019) shows that customers do not have a complete understanding of the level of environmental obligations of the hotel, which was not observed in the results of this study and the analysis of Iranian tourists.

**Conclusion**

According to the results of the study, it is suggested to marketing managers, especially in the field of tourism and hotel management to pay high attention to the components identified in this study as key indicators of customer satisfaction and gaining a competitive advantage and to use all the forces and capital of their organization in order to realize the value-creating components and prevent the barriers to value creation. Managers and entrepreneurs can use the big data to check the importance and validity of their new ideas and use it to identify and predict possible customer feedback. Also, due to value being relative from the customers’ point of view, organizations should measure and present the values expected by their business customers, and more specifically, the desired values of their target market, and avoid looking at customer values in general. Due to the rapid changes in customer demands and expectations, these organizations should always and periodically evaluate these values in their target market to avoid changing customer expectations. It is recommended that they use social media data to validate their ideas, attitudes, and analyzes to identify existing weaknesses and provide new services. Business managers are advised to focus on the most important value dimensions that can be developed by examining their strengths and weaknesses, if they do not have the necessary strength and capital to achieve all the values expected by customers and to create a competitive advantage for themselves in that field. In other words, to gain a corner of the
market by focusing on that segment. Organizational marketing and R&D researchers are also encouraged to make extensive use of artificial intelligence processes, machine learning approaches, and big data analysis in their research and customer-related fields; Also, they should always be up-to-date and know the latest technologies and approaches, learn and use applied approaches in this field so with higher accuracy, faster examinations, a bigger scope of research, reduced costs, they can present optimal goods and service in accordance with the demand and expectations and take steps to improve the performance of the marketing unit and the organization as a whole.

The present study is based on the opinions of customers in the Iranian tourism industry and tourist areas of Tehran province and researchers are recommended to conduct this research globally and compare the results. Also, it is suggested that by combining classical methods of analyzing and predicting consumer behavior with new methods of data science such as machine learning and deep learning, they use the potential capabilities of big data analysis in market research and other branches related to business management.

Regarding the limitations of the research, we can mention the use of non-supervisory method for clustering comments in the research method. To solve this problem, researchers are suggested to process the text and compare the results by combining two methods of supervised and unsupervised. Also, the limitation of the statistical population due to the relative lack of comments in the field of Iranian tourism on the Internet, creates the possibility of sampling errors.

Acknowledgement:
The authors are thankful to the tourists of Iran and Tehran province, for collaborating with this study, for data collection.

Funding:
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest:
The authors have no conflicts of interest to declare.

References
Information. In Marketing Management (pp. 177-235). Springer, Cham.


Faezi Rad, M., & Pouya, A. (2016). Clustering of online stores from the supplier's point of view with the help of optimizing the number of clusters in the two-stage SOM algorithm. Iranian journal of Industrial Management Studies, 14(43), 109-134.


