Application of data mining techniques to present model of tourist’s health behavior analysis

Masoumeh Sheikhhassan¹, Mehdi Sadeghzadeh²*, Mahdi Faghihi³

¹ Student of information technology management, department of management, faculty of management and economics, Tehran Science and research branch, Islamic Azad University, Tehran, Iran.
² Assistant professor of computer engineering, department of computer, faculty of electrical and computer engineering, Mahshahr branch, Islamic Azad University, Mahshahr, Iran
³ Assistant professor of information technology management, department of management, faculty of management and economics, Tehran Science and research branch, Islamic Azad University, Tehran, Iran.

*Corresponding authors: Mehdi Sadeghzadeh, Assistant professor of computer engineering, department of computer, faculty of electrical and computer engineering, Mahshahr branch, Islamic Azad University, Mahshahr, Iran. Email: m.sadeghzadeh@mhriau.ac.ir

Received 2021 January 31; Accepted 2021 March 07.

Abstract

Background and Objectives: Today, the increasing desire of consumers for health tourism has led to a greater understanding of the behavioral patterns of tourists. It becomes clear that intervention in that process is necessary to achieve the desired results. The development of tourism services at a specific time is essential for the target market and meeting the needs of tourists to succeed in the tourism market.

Methodology: In this study, both quantitative and qualitative methods were used. In the qualitative method, 50 cases were interviewed, and in the quantitative part, 156 questionnaires were distributed, and finally, its validity and reliability were examined. SmartPLS software (version 2) was used for modeling and data analysis.

Results: The effective factors in tourists’ decision to choose Iran as a health tourism destination were examined after analyzing the data. Based on the result, the needs of health tourists can be met and more motivation can be created by focusing on the obtained factors.

Conclusion: In this article, the behavior of health tourists was analyzed, and finally, a model based on tourist behavior was designed to better manage capacity and meet the challenges and needs of tourists. As a result, agencies can predict their future behavior based on the past behavior of tourists.

Keywords: Medical tourism, Health behaviors, Cluster analysis, Data mining

Introduction

Today, the marketing process is becoming more complex and sensitive every day with the increasing growth of information technology, suitable communication platforms and intelligent technologies, the competitiveness of the tourism market, and the growth of the number and information of customers. Furthermore, paying attention to the issue of relationship with the customer, management is becoming more important every day and the budget allocated to it is increasing in organizations. Leading organizations always try to design new products according to their customers’ opinions and offer customized goods according to the required needs of customers, which motivates them to travel to the selected destination (1).

Companies and tourism service providers have a lot of data about their customers, which provided that it is stored in an organized and integrated manner in the database, it can be used for marketing decisions. The identification of important variables and their relationships is often a difficult task that can be improved with the help of data mining. This will increase visitors, retain tourists and attract investors, and improve Iran’s economy. Various types of research have been conducted on the tourism industry in this domain. One of the requirements of this
research is that due to the complex behavior of the new generation of health tourists and the increasing demand for different products, it has become very difficult to understand their behavior. Accordingly, this research presents a model based on the analysis of the behavior of health tourists in which the relationship among such variables as infrastructure, destination image, motivation, information resources, and tourist behavior is examined.

Theoretical foundations and review of past studies
Organizations are trying to gain a greater share by understanding the needs and behavior of tourists in competitive environments. The tourist experience of a destination is a fundamental product in tourism. For this reason, the tourist destination is in the center of the competition. The features of a tourist destination, including natural resources, cultural resources, human resources, facilities, medical services, and other activities, should be limited to the tourist visit and offered to others to visit the destination. Consequently, it is not enough for the destinations to be in the minds of the customers but it must be so unique and distinctive that be selected as the final decision of the tourists. The mental image can represent a simplified set of fragmented perceptions that contain a lot of information about a place.

The destination image also identifies the strengths and weaknesses perceived by the tourist. In this regard, E-tourism helps tourists to have a targeted program led by the government, travel agencies, and centers in the shortest time, with the least facilities and costs to identify the tourism capacity of a country. In this regard, providing a comprehensive tourism system, developing channels and public educational space for tourists to facilitate the process of booking, choosing a tourist destination, and using historical, natural, and environmental attractions in one step without frequent visits to offices and organizations are very impressive.

Advertisement display on channels is carried out both offline and online. In offline advertising, which is an individual’s experience from his previous trip, the level of his satisfaction from the trip is announced to another person, while in online advertising, information is received virtually through social networks and sites. As a result of improving information and communication, providing facilities and discounts have taken a significant role in advertising. Researchers have also developed various patterns to motivate health tourists to travel to Iran. This paper presents a model based on the analysis of the behavior of health tourists.

Background research
In a piece of research, the behavioral pattern of tourists has been studied and identified, the laboratory model of the destination management system has been implemented, and the appropriate response to passenger demand by data mining techniques has been presented. A study predicts online customer behavior, and in other researches, the behavior of tourists has been extracted using the data mining method in the telecommunication company, in which the clustering method has been used to analyze customer behavior. The existence of inconsistencies in tourism marketing strategies is examined in another study.

This research aimed to, firstly, identify discourse differences in expressing the position of regional tourism marketing strategy documents and, secondly, prioritize the identity of a specific destination. In a study, a forecasting framework has been provided that uses machine learning and Internet search indicators to predict tourist arrivals. Accurate forecasting of tourist numbers can help tourism professionals optimize and manage capacity, and research has identified components and content analysis for medical tourism facilitator websites. In another piece of research, the Apriori algorithm has been investigated, which has a serious weakness in time calculations, and due to the repeated searches, it has to scan the database repeatedly for each combination, which is very time-consuming. The present study deals with the effect of clustering algorithm on Apriori algorithm that the combination of these two algorithms produces more information produced in the shortest time.

Research method
The first part of the research is described using a qualitative method and the second part is described using a quantitative method.

In the qualitative method, the researcher first designed the interview questions by reviewing the literature and theoretical frameworks. The electronic version was designed in Google Drive and its link was shared. Sampling was continued gradually with 50 cases until reaching data saturation. The samples were selected using the maximum variation sampling method and the sampling continued until new findings were obtained from the new textual data. Part of the data was given to a researcher who had nothing to do with the study and was an external observer. To determine if he had a similar understanding of the data and also to increase the reliability, the external arbitrator was used and its validity was examined.

In another part, to measure the research variables, a questionnaire was designed based on the results of content analysis, which included 38 items in the field of health tourism. This tool was designed based on the Likert scale and the participants, who were part of the same qualitative section, were provided with the link of the questionnaire, and 156 questionnaires were used to analyze the data. To validate the questionnaire, the experts’ opinions were applied, and to model the conceptual model of the research, the method of structural equations with SmartPLS software was used.

Research Findings
One way to calculate reliability is to use Cronbach’s alpha formula. If Cronbach’s alpha value is greater than 0.7, it indicates the reliability of the questionnaire items. Given that the mean-variance and Cronbach’s alpha coefficient for all research variables were higher than 0.5 and 0.7, respectively, it can be assumed that the research tool had an acceptable convergence validity. Moreover, to examine the divergent validity, firstly, the correlation matrix of the main components of the research model was created.

By placing the square root of average variance extracted with the value of 1 in the original diameter of the matrix in Table 1, it was observed that this value for each variable was greater than the correlation of one structure with other structures. As a result, the research tool had good divergence validity.

According to the confirmation of the desirability of the factor loading values, the convergence validity and the divergence validity of the research measurement model fitness were approved.

The second steps in procedures were the use of path analysis, coefficient of determination, and model fit index.
In path analysis, the relationships between variables flow in one direction and are considered separate paths. The concepts of path analysis are best explained through its main feature (i.e., the path diagram), which shows possible causal links between variables (Homan, 2008). Figures 1 and 2 depict the structural equation modeling and the path diagram along with the significance coefficient and path coefficients.

**Research hypotheses testing**

In this section, the research hypothesis was examined and tested, in which the significant levels of the model paths indicate whether the research hypothesis was significant or not. Furthermore, standard path coefficients show what percentage of changes in the dependent variable is explained by the independent variable. In this regard, if the majority of the numbers are higher than 1.96, the importance of the path between the two variables can be confirmed and this relationship can be proved. Therefore, all research hypotheses are accepted.

The section of the sub-hypothesis test refers to the effects of all variables on each other, including direct and indirect effects. These effects have been calculated by SmartPLS software (version 2) and tabulated in Table 3.
the above table, all research hypotheses were accepted, with the existing infrastructure and potential as the highest impact (0.257).

**Conclusion**

According to the proposed model, it can be concluded that the variables of motivation, infrastructure, channels, and information resources played a significant role in choosing a destination. Furthermore, the appropriate information channels and resources and infrastructure of the destination city could motivate health seeker tourists to cure the disease. It was also revealed that all model variables affected the behavior of health tourists. As a result, it is recommended to make short films about health tourists on Iranian radio and television and share them on social networks, the measure that has an important role in motivating, and ultimately, attracting health tourists. Additionally, it is suggested to have a suitable platform to create powerful teams to respond to word-of-mouth and offline electronic advertisements of tourists on the
sites, hold training courses for foreign tourists, and design various advertising programs for tourism industry intermediaries to introduce travel packages by the Iran Cultural Heritage, Handicrafts, and Tourism Organization and smart campaigns. It is recommended to perform future research to analyze the behavior of tourists using variables, such as technology, cost, revenue, and principles, as well as laws of different countries.

The results of a piece of research showed that motivation, communication activities, information resources, and infrastructure and potentials had a significant impact on the choice of the destination image (14). This finding was consistent with those of research carried out by Minaei (15). Channels and information sources are very effective in advertising and motivating tourists, leading to attract more tourists and buy more services by them, and eventually, more profitability. For this reason, most tourists, based on the experience and mentality of their previous trip, offer their visit experience to others using word-of-mouth and offline advertising.

The results of this study are in line with those reported in a study carried out by Taghipourian et al. (16) regarding the dimensions of the proposal and the visit in the variable of the destination image. Tahmasebi and Roshanian (17) reported that interfering factors, including social, economic, psychological, cultural, and environmental factors, are influential factors affect the variable of tourism behavior, which is consistent with the results in this study and that in research performed by Slabert regarding the variable of motivation (18). Nevertheless, considering other infrastructural factors, destination image, and information sources, it was not in agreement with the findings of the present study. Furthermore, presenting different travel suggestions according to the different behaviors of tourists is consistent with Bagheri and Hashemzadeh (19, 20) in creating tourism recommendation systems.

**Recommendations and restrictions**

One of the limitations of this research was related to the lack of access to databases of all sectors due to the expansion of the tourism industry. The other limitation was the lack of sufficient theoretical and experimental background on the research domain. Other limitations of this research included advertising, media, and educational limitations due to the absence of appropriate infrastructure and platform. Infrastructural, cultural, and religious constraints, as well as the concentration of industries and services in several large cities, were other limitations of this study.

**Acknowledgment**

The authors would like to express their gratitude to the general office of supervision and accreditation of the Deputy Minister of Treatment and the manager of the Health Tourism Office at the Ministry of Health and Medical Education.

**References**

5. Breslin J. G, Passant & Decker S. Introduction to the social web (web 2.0, social media, social software). In the social semantic web. Springer 2009; pages 21-44.
10. Jeuring, Discursive contradictions in regional tourism marketing strategies: The case of Fryslân the Netherlands1. 2017; Perspectives on proximity tourism in Fryslân.