RESEARCH ARTICLE | MARCH 07 2024
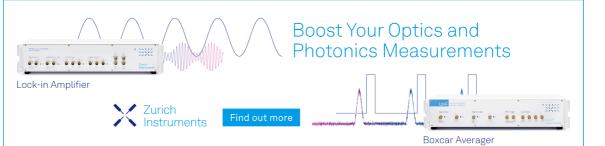
The implementation of data mining technology in tourism industry

Siti Aishah @ Tsamienah Taib ✉ ; Noratikah Abu; Azlyna Senawi

Check for updates

View Online   Export Citation   CrossMark

08 March 2024 09:07:53

The Implementation of Data Mining Technology in Tourism Industry

Siti Aishah @ Tsamienah Taib[a)], Noratikah Abu[b)] and Azlyna Senawi[c)]

Centre for Mathematical Sciences, Universiti Malaysia Pahang, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia

[a)]*Corresponding author: tsamienah@gmail.com*
[b)]*atikahabu@ump.edu.my*
[c)]*azlyna@ump.edu.my*

Abstract. Data analytic approach has been widely used in combining with the traditional method in order to improve the accuracy of forecasting, especially in tourism industry. Currently, the common used approach is known as data mining technology, which can dramatically accelerate the recovery process of the tourism industry after the outbreak of COVID-19. The purpose of this paper is to review the implementation of data mining technology in tourism industry based on different case studies that were collected through selected digital databases. Based on the literature reviews, there is no doubt that data mining technology has the potential to be extremely important in tourism research. This technology can help the industrial practitioners make wise decision in future since it provides an efficient approach for identifying and comprehending customer behavior's pattern. It is hoped that this study can contribute as a reference for the future studies and are able to help destinations and tourism businesses maintain continuous supplies of tourism products and services in order to fulfill the customer's demand.

INTRODUCTION

Nowadays, we are greatly affected by the COVID-19 outbreak. Since the number of infections and deaths has dramatically increased, the World Health Organization has declared this outbreak as a global pandemic. Therefore, to curb the spread of the COVID-19 epidemic, the government had to implement several restrictions such as school closures, movement controls as well as domestic and international tourism controls. Due to the closure of borders and travel bans (internal and external), the tourism industry has been greatly affected and caused enormous losses. Based on the UNWTO record, in 2020 the percentage of tourist arrivals (in international tourism receipts) decreased by 20% to 30% (300 to 450 US$ billion) [1]. As the tourism sector is the largest contributor to the economic development and stability of a country, the government takes seriously the need to rehabilitate the sector.

Tourism demand forecasting is very important for a lot of organizations such as government bodies, researchers, policy makers and industrial practitioners since prediction of future events is one of the crucial pieces of information that strongly influences the future decision making process. Based on [2], future decision making are depends on the accurate forecasting. The right decision can help the industrial practitioners to forecast the infrastructure capacity needed to accommodate tourists in the future. Furthermore, the government bodies are able to project revenues obtained from the tourism industry [3].

The formal statistical approach was introduced in the tourism industry since appropriate analysis is needed to make the right decision. But, due to some limitations of statistical methods, the researchers decided to apply data mining technology for tourism data analysis. The common uses of data mining technology in the tourism industry are as follows: (1) forecasting arrivals of tourists to a location, (2) forecasting their total expenditures during travelling and (3) analyzing profiles of tourists [4].

Based on the literature, until present, researchers still study on the implementation of modern technologies in tourism industry, but we are attempts to review the data mining technology in tourism industry. Therefore, the aim

of this paper is to review the implementation of data mining technology in tourism industry based on different case studies.

DATA MINING FOR TOURISM INDUSTRY

Data Mining Technology

Nowadays, data analytic approach has been used in combining with the traditional model as it has advantage in improving the forecast accuracy [5, 6]. Most of researchers interested to use data mining technology in forecasting the future tourism demand, which can dramatically accelerate the recovery process of tourism industry. Data mining techniques or knowledge discovery also known as computer-assisted process to discover patterns and hidden relationships in a large data set. Knowledge are discovered and presented in a form that is easily comprehensible to humans by using machine learning, statistical, mathematical, visualization and database [7]. Data mining process is divided into two parts: data pre-processing and data mining. Data pre-processing involving data cleaning, data integration, data selection and data transformation [8, 9].

According to [10, 11], there are various data mining techniques that can be used in forecasting such as association, classification, clustering, decision tree, neural networks and etc. In the following section, those data mining techniques are outlined briefly.

Association

According to [10], association is one of the outperform data mining techniques and it is used to identify the trend between a variable with other variables in the same activity.

Classification

Large data sets were divided into different groups or classes by using this technique to produce more accurate prediction. To define a certain group or class, classification may be used to develop the category of a customer, object, or item in a data collection by specifying several attributes [11]. In addition, classification can be used together with other approaches, such as decision trees to identify classification or clustering to locate groups of similar qualities in distinct categories.

Clustering

Clustering was one of the first data mining techniques employed. In order to comprehend the differences and similarities between the databases, the clustering method comprises an analysis of one or more qualities to discover data that are similar to each other. The clustering method is also known as segmentation since it divides data into categories in order to locate a cluster of related outcomes [11].

Decision tree

Based on the values of certain input variables that have the highest connection with the target variable, this approach divides data into naturally exclusive and exhaustive subgroups. The final output may be represented graphically by a tree-like structure, which provides a concise description of the data [12].

Neural networks

Neural networks considered as most powerful data mining technique for forecasting due to its ability to represent complex relationship in data well. Originally, this approach was designed to mimic the human brain, consisting of a number of interconnected processing nodes or known as neurons. Each neuron will transmit the input signal to the other neurons, and produces a transformed output signal to other neurons [11, 12].

Methodology

Literatures regarding the implementation of data mining technology in tourism industry were collected through selected digital databases such as Scopus, Springer, Science Direct, IEEE Xplore and Web of Science. We use other related words regarding tourism research rather than the keyword tourism, tourist, forecasting and data mining. The logical OR, AND and NOT descriptors have been used also during the literature search. Then, the primary studies were selected manually in the screening process.

Applications of Data Mining in Tourism Industry

In many industries, data mining technology is the most significant approach for predicting. In today's ever-changing economic climate, we always have plenty of opportunities to apply numerous strategies to make business decisions. Since data mining technology has advanced dramatically nowadays, there is tremendous value in the interdisciplinary concept of data mining for forecasting to tackle challenges in the tourist business. There are a lot of techniques in data mining technology that may be used for various challenges and give insights on those specific future business problems. Therefore, it is highly recommended to identify the business problems first before applying any data mining techniques in order to execute the best results. According to keyword-driven search and content analysis, eight articles have been identified as relevant and will be elaborated on further in this section.

In 2017, [13] introduced data mining knowledge and use Apriori Algorithms, which is one-dimensional association rule to generate reliable and useful tourism information recommendation that can help tourists make a right decision. The data set used mainly focuses on the tourists' individual hobbies and interest. The findings showed that data mining techniques will give a huge contribution in the future tourism industry.

In addition, [8] discovered the relationship between tourists' demographic information and their online transactions by applying data mining method together with association rule techniques. In this paper, the researchers' aim is to determine a tourist's interest based on search behavior when they search for tourism destination. Since information technology (IT) have been widely integrated in tourism industry, therefore this study focuses on tourists who are using mobile devices to search for travel destination options such as lodging, tourist hot spots, activities, restaurants and gift shops. Their findings demonstrated that the industrial practitioners can effectively accommodate the tourists' demand by applying data mining method in tourism industry.

A large amount of information regarding tourism attractions is displayed on travel portal websites nowadays. Sometimes, the information provided led the users to make the wrong decision. Therefore, [14] proposed a tourism recommendation system based on data mining incorporated with filtering algorithm to solve the problem. Experimental findings indicated that the collaboration between data mining and filtering algorithm are able to produce a good tourism recommendation system that can reduce the risk of decision making in future.

[15] designed an Android Smartphone application in tourist guide system for Mumbai City. The researchers recommended K-Nearest Neighbor (KNN) Algorithm used in the system since it can identify the nearest locations with the highest accuracy. Based on the results, [15] concluded that the details about a place can be provided directly to the tourists through this system and at the same time the tourists will not miss any interesting places to visit in Mumbai City.

Most researchers focus on smart tourism or e-tourism to improve the efficiency of tourism management. [16] developed a smart tourism information system using big data technology, and conventional K clustering techniques were incorporated to enhance the performance of the final output. Based on the experimental results, this new system constructed together with data mining technique can provide tourism information needs and user experience needs. Meanwhile, [17] applied Recency, Frequency and Monetary (RFM) analysis, one of data mining techniques for identifying the best way in profiling the customers by using actual data obtained from the customer relationship management system (CRM). K-means algorithm was implemented in this study to cluster 369 profitable hotel customers into eight clusters. [17] then conclude that this analysis may prompt the top management to devise new approaches to improve the efficiency of CRM system.

There were also research that incorporated data mining in agro-tourism industry. [18] established association rule for arranging and clustering the activities of agro-tourism in orchard, and claimed that reaping and tasting fruit activities must be included in the list of activities conducted in each orchard.

Last but not least, [19] used Apriori Algorithm to determine the trend of tourism visit to Bali in the period 1982 to 2018. Based on the calculations using Weka Data Mining application, they found out the trend of tourists visiting Bali is association rule with the best confidence values.

Forecasting Accuracy with Data Mining Technology

A lot of work done before regarding the usage of data mining technology in forecasting. For each case study, the researchers will select several appropriate data mining techniques and conduct an experiment to identify the best technique. Seven papers were found in this category and are outlined briefly in this section.

In the first paper, [20] tested the forecasting accuracy of four forecasting techniques by using the series data set from the foreign tourists' arrival in the district of Central Java's, which comprised Magelang, Solo and Wonosobo. In order to achieve the goal, monthly data from 1991 to 2013 were used to evaluate the accuracy of the selected models. These four forecasting techniques used are Backpropagation Neural Network (BPNN), K-Nearest Neighbor (KNN), Multiple Linear Regression (MLR) and BPNN with Genetic Algorithm optimization (BPNN-GA). GA is employed because there are certain parameters (learning rate, training cycle and momentum) that need to be optimized. Root Mean Square Error (RMSE) is used to measure the performance of these four forecasting techniques. From the observations, they found that the best forecasting technique among candidates technique was BPNN-GA since its value of error were the lowest. The Genetic Algorithm is proven to be capable of optimizing the parameters of BPNN-GA while also minimizing the inaccuracy of BPNN's prediction.

In 2016, [21] implemented multiple linear regression, multilayer perceptron (MLP) regression, and support vector regression (SVR) to make multivariate tourism forecasting for Turkey. The data set used is monthly arrival information of tourists from January 1996 to December 2013. By using three forecasting accuracy measurements: Root Relative Squared Error (RRSE), Relative Absolute Error (RAE) and correlation coefficient (R), these models were measured and compared through the same validation data. The researchers claimed that SVR model outperforms the other models for tourism demand in Turkey since it showed accuracy with R = 0.9901, RAE = 12.34%, and RRSE = 14.02%.

[22] performed eight forecasting models in order to determine the best estimation model of tourism demand of Spanish tourists who visit Cappadocia, the most beautiful tourism regions in the world. These models are Artificial Neural Network model, Naive model, Moving Average, Exponential Smoothing, Multiple Regression, Winter's Exponential Smoothing Method, Box-Jenkins Model (ARIMA) and Holt Exponential Smoothing Method. According to the analyses, Artificial Neural Network model and Multiple Regression model were appointed as the best models based on the forecasting accuracy examined.

Moreover, [23] compared five algorithms of three data mining classification techniques in order to identify the outperform algorithms for tourism knowledge discovery. Those five algorithms are listed as follows:

TABLE 1. Five Algorithms of Three Data Mining Classification Techniques

Category	Algorithms
Decision Tree	C4.5 (J48 in Weka)
	Random Forest
	Projective Adaptive Resonance Theory (PART)
Artificial Neural Network	Multilayer Perceptron (MLP) model
Support Vector Machine	Sequential Minimal Optimization (SMO in Weka)

All algorithms were implemented and the best performance is recorded before and after attribute selection. Their results showed that Random Forest outperformed all other algorithms before attribute selection meanwhile J48 (C4.5) was considered as the best algorithm after attribute selection. In 2021, [24] also compared four data mining classification techniques, which are JRIP, Random Tree, J48 and REP Tree to discover the most effective algorithm that can be implemented to boost the business opportunities in travel-related fields. The implementation of these four data mining classification techniques was carried out by using Weka application. Based on the results, they found out J48 outperforms the other algorithms, attaining the highest forecasting accuracy.

[25] demonstrated Multiple Linear Regression (MLR) in Artificial Neural Network (ANN) to forecast the number of tourist arrivals in Indonesia. In this study, MLR acts as a correlation test predictor variable while ANN as a forecasting tool, and the researchers applied back-propagation algorithm in the network training and testing process. Both methods work well together since MLR method produce strong correlation between the variables and ANN method is able to provide the accurate prediction with a percentage Error Value of 0.089% and an Accuracy Value of 99.9911%.

[26] analyzed and compared four algorithms to identify the optimal data mining method for small-sized hotels. These four algorithms are M5P Algorithm, M5 Rule Algorithm, JRip Rule Algorithm and J48 Algorithm. They

concluded that M5P and M5 are convenience in tourism demand forecasting while JRip and J48 are suitable for analysis of guest behavior.

CONCLUSION

Even though the history of data mining is short, but data mining technology emerges parallel with the time require, in response to the rapid development of information technology and the need to make judgments based on complex data set. Most of the researchers apply data mining techniques such as association rules, clustering, decision tree, artificial neural network etc. in forecasting across various industries. Based on the literature reviews, there is no argument that data mining has the potential to be extremely important in travel and tourism research. Data mining can help the industrial practitioners make wise decision in future since it provides an efficient approach for identifying and comprehending customer behavior's patterns. To summaries, data mining technology is a must in tourism industry in order to acquire more useful and valuable information to improve the effectiveness of management. In addition, tourism industry needs this technology to progressively improve the tourism destination's competitiveness. Lastly, this study can contribute as a reference for the future studies and are able to help destinations and tourism businesses maintain continuous supplies of tourism products and services in order to satisfy the increasing demand for travel experiences.

ACKNOWLEDGMENTS

REFERENCES

1. UNWTO (2020).
2. D.C. Frechtling, *Forecasting Tourism Demand: Methods and Strategies*, (Butterworth-Heinemann, Oxford, 2001).
3. S.F. Witt and C.A. Witt, Int. J. Forecast. **11**, 447-475 (1995).
4. I. Bose, *Encyclopedia of Information Science and Technology*, Second Edition, (IGI Global, USA, 2009), pp. 936-940.
5. T. Sutthison, NUJST. **27**, 18-31 (2019).
6. S. Xu, H.K. Chan and T. Zhang, Transp. Res. E: Logist. Transp. Rev. **122**, 169-180 (2019).
7. P. Cabena, P. Hadjinian, R. Stadler, J. Verhees and A. Zanasi, *Discovery data mining: From concept to implementation*, (Prentice Hall, New Jersey, 1997).
8. P. Juwattanasamran, S. Supattranuwong and S. Sinthupinyo, Int. J. Eng. Sci. **2**, 38-44 (2013).
9. G.K. Sharma and P. Sharma, IJLTET. **7**, 580-587 (2016).
10. K.M. Raval, IJARCSSE. **2**, 439-442 (2012).
11. A.S. Osman, IJDSR. **2**, (2019).
12. M.B. Karathiya, R.S. Sakshi, D.S. Sakshi and D.R. Kathiriya, J Inform Oper Manag. **3**, 114-118 (2012).
13. W. Zhu, Adv. Eng. Res. **123**, 93-98 (2017).
14. Z. Wang, B. Liu and B. Liu, J. Phys. Conf. Ser. **1345**, (2019).
15. P. Yawalkar, S. Mahajan, D. Dhage and P.V. Naval, IJARIIE. **8**, 1403-1406 (2022).
16. J. Du, Mob. Inf. Syst. **2021**, (2021).
17. A. Dursun and M. Caber, Tour. Manag. Perspect. **18**, 153-160 (2016).
18. K. Pitchayadejanant and P. Nakpathom, KJSS. **39**, 407-413 (2018).
19. D.A.K. Arimbawa, N.A. Prabawa and P.A. Mertasana, IJEET. **4**, 10-14 (2019).
20. E. Noersasongko, F.T. Julfia, A. Syukur, Purwanto, R.A. Pramunendar and C. Supriyanto, Indian J. Sci. Technol. **9**, (2016).
21. S. Cankurt and A. Subasi, Turk. J. Electr. Eng. Comput. Sci. **24**, 3388-3404 (2016).
22. E. Islamoglu and N.O. Dogan, PAUSBED. **40**, 189-208 (2020).
23. T. Urgessa, W. Maeng and J.S. Lee, World Acad Sci Eng Technol. **11**, 119-124 (2017).
24. C. Miao and T.S. An, Baghdad Sci. J. **18**, 737-745, (2021).

25. R. Sovia, M. Yanto and Yuhandri, TURCOMAT. **12**, 1492-1501 (2021).
26. A. Njegus, V. Nikolic and V. Jovanovic, "The selection of optimal data mining method for small-sized hotels," in *International Scientific Conference of IT and Business-Related Research*, (Synthesis, Serbia, 2015), pp. 519-524.

08 March 2024 09:07:53