International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

Use of Data Mining in Enhancement of Tourism

Mahvish Fatima*, JitendraKurmi
Center of Engineering and Technology, BabasahebBhimraoAmbedkar University, Lucknow, Uttar Pradesh, India

ABSTRACT

Data Mining is the procedure of discovering new, raw and interesting patterns from the previously based data repositories. It is entitled as to determine useful information with the help of different algorithms. Data Mining is applicable to almost every field presently for example banking, hospital, market basket analysis, education, CRM, fraud detection, tourism etc. Tourism is the key element in the economy of any country. About 12% of the country's economy comes from the tourism. It is the sector, which is for people and from people. A number of experiments are going on in the field of data mining related to tourism but if we compare with other sectors, it is still at the early stage of development. Therefore, there is a need to focus more on the tourism sector from the research perspective. This paper elucidates the use and work of data mining in the tourism sector up to date more effectively.

Keywords: Architecture, Data Mining, Clustering, Regression, Review

I. INTRODUCTION

TOURISM is the important contribution to the economy of a country. It is contributing approximately 12% to the GDP of India. It is growing larger and occupying a big space in the area of investment. Generating more insights about the tourism that how effective it will be in the future certain areas of technology is working on it nowadays. One such area dealing the tourism is Data Mining.

India is the 7th largest tourism economy in terms of GDP in the world says WTTC [11]. Moreover, in the upcoming time, it is likely to increase from 14% to 17%. Data Mining refers to the art of extracting meaningful information from the previously known data sets to give birth to some new and interesting facts. It is of two types Predictive and Descriptive. It is being use currently in almost every field but its application to the field of tourism is still new. There are several techniques of data mining are available which is being used. Tourism is becoming one of the most significant aspects of the world's economy.

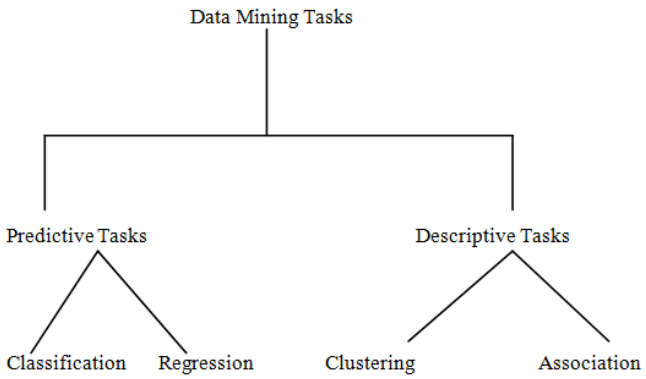

Figure 1. Data Mining

Data Mining helps to extract the unknown patterns from a known one. It is also termed as Knowledge Database Discovery (K.D.D.). Generally, data mining has these steps in the extraction:

1) Data Cleaning
2) Data Integration
3) Data Selection
4) Data Transformation
5) Data Mining
6) Pattern Evaluation

7) Knowledge Presentation

Several traditional and statistical techniques are being use for predicting the tourism. By applying data mining in the field of tourism, we can progress more in terms of economy. The main aim of this paper is to highlight the methodologies of data mining needed these days for improving and predicting the tourism of a country. Finally, this paper helps in recognizing some currently used data mining methodologies.

II. INTRODUCTION TO RETROSPECTIVE APPROACH

The contribution of tourism in the economy of a country is inevitable. The researchers, planners, and forecasting team are working uninterrupted to gain positive results. Primarily researchers have used statistical or econometric techniques in the past for predicting the market performance of future. However, these techniques provided only a slight improvement despite the major efforts applied. Due to this time taken concept scientist started to apply data mining to tourism demand forecasting and have achieved fruitful results [1].

Time Series methods of forecasting generate the future values based on the variable's past values while Casual Methods of forecasting describes the relationship between the independent and dependent variable. However, if we will talk about the advancement of data mining in tourism then it is still at its developing stage. There is more research needs to be done in order to find some new insights into the field of tourism demand forecasting through data mining.

III. ARCHITECTURE OF DATA MINING

Data mining comprises of several components in its architecture. When the data set is processed, it goes through all these components, which are describe as below: [14]

1) Database
2) Database Server
3) Knowledge Base
4) Data Mining Engine (functions)
5) Pattern Evaluation Model
6) Graphical User Interface.

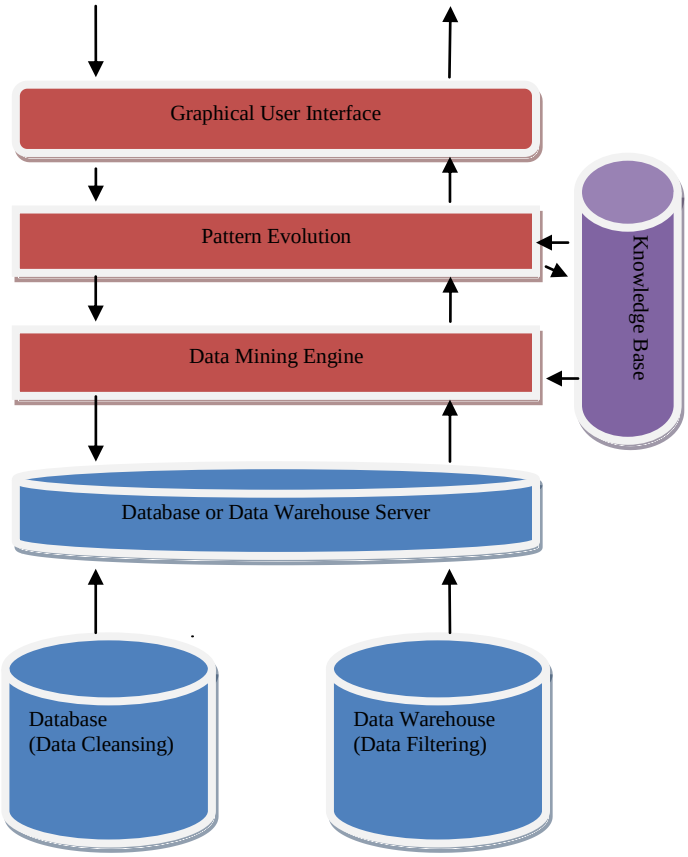

Figure 2. Architecture

A. Database

No data is generated for data mining; it is collected from information repositories, spreadsheets and other documents, which contain a large number of data. Data is usually stored in the databases or data warehouses. The Database is the main source of information for data mining.

B. Database Server

The server contains the information, which is retrieve from the database or the data warehouse. Its work is to generate the data for the mining process.

C. Knowledge Base

This is useful in the process of data mining to compare the results of mining with the information in the database. It contains a set of user experiences

so sometimes it also provides input to the data mining engine. Thus, it works as a guide in the process of data mining to tell what is right and what is wrong.

D. Data Mining Engine

The Engine is the core component of any system. In the same way, data mining engine is also the core part of the architecture of data mining. It has a number of modules for executing data mining tasks such as classification, clustering, regression, time series, etc. [14]

E. Pattern Evaluation Module

It estimates the interestingness of the pattern evaluated by the data mining engine. A threshold value is there from which they obtained value is evaluated for finding the interesting patterns.

F. Graphical User Interface

G.U.I. acts a mediator between the data mining system and the user. It helps the user to communicate with the system effectively and easily. The user only gets to see the basic information needed not the complexity behind it. [13]

So, these are the components involved in the process of data mining.

IV. DATA MINING TECHNIQUES

Data Mining algorithm is a set of calculations and heuristics that designs a new model from the old data sets available. For making a data mining model first, it analyses the system, then checks for the patterns and finally generates a result of the new model. [7]

Performance of algorithm is measured in two ways:

By time or

By space

There are many types of techniques available in the market for performing data mining some of them are:

1) Classification
2) Clustering
3) Regression
4) Association
5) Prediction
6) Decision Trees
7) Visualization

A. Classification

Classification is the process of labeling unknown class models that distinguishes from other class models in the data set. It can be represented as Classification Rules (i.e. IF…THEN rules), Neural Networks, Decision Trees or Mathematical Formulae. [7] For example, we can easily classify it according to the type of items. Hence, the classification predicts the discrete and unordered labels.

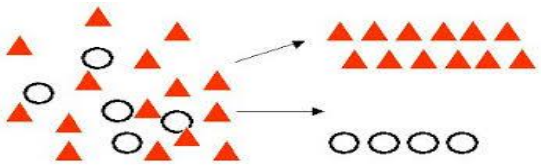

Figure 3 Classification

B. Clustering

Clustering explains about how the similar patterns are clustered together and different ones are separated from the group to make a new cluster. It is the process of combining the same type of object from different clusters into a new one. It is useful for an organization or company in making business strategies which gives profit to the company. It is of two types Central and Hierarchal.

C. Regression

It is the technique of predicting any situation related to energy demand, weather forecasting etc. It helps us to find out bond the between variables. There are generally two types of regression:

1) Simple or Linear Regression
2) Multiple Regression

In Simple Regression analysis is based on a single explanatory variable.

$$E(Y|X) = \beta_0 + \beta_1 X$$

In Multiple Regression analysis is based on multiple explanatory variables.

$$Y = a + b_1x_1 + b_2x_2 + ... + b_kx_k$$

It usually comes under predictive modeling.

D. Association

This technique correlates between two items. They represent the difference between two items. For example, the association between humans and animals is that they both have a life i.e. they are living beings. The two mostly used techniques for association are Apriori algorithm and FP-Growth algorithm. However FP-Growth generates much better results than apriori. [9] One of the examples of association is market-basket.

E. Prediction

Prediction means to give a statement what will happen next in prior to previous knowledge and experience. It is also useful in recognizing the theft and in prediction of profits too. It is use for determining latest trends and work accordingly.

F. Decision Trees

Decision Tree is the algorithm for classification and regression models. It can work on both types of data either categorical or numerical data. The following decision tree shows a vehicle buyer according to their gender.

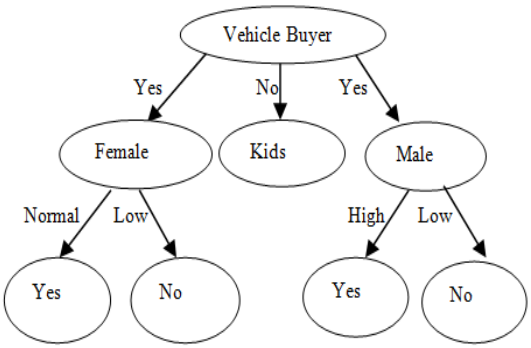

Figure 4. Decision Tree

G. Visualization

This technique helps to provide the graphical approach to the results obtained from data mining. It helps in summarization of data and reporting of results. Due to its color combination, it can provide a much better and enhanced view of data mining activities. It gives a thorough understanding of the data sets to the user. [2]

V. FINDINGS AND RELATED WORK

According to Rob et.al. [1] Data Mining techniques are still in its developing phase in tourism, comparable to other sectors. Manoj et.al. [2] stated the use of data mining in the field of tourism and also elaborated the negative points of data mining which still needs improvement. Pairaya et.al. [3] incorporated the travelling patterns of the customers in his research. They have used the search pattern of customers to analyse the travelling choices. This helped the tourism sector to know the traveller's demand effectively. Mirjana et.al. [4] reviewed a large number of papers approximately 88 that analysed the applications of data mining in tourism sector. It concluded that tourism profitability not only depends on data but on people choice too. Later Prabal et.al [5].introduced the concept of gamification to the data mining. He worked on it to gain insights about the data mining impacts on the tourism. Tourism industry not only depends upon the travelling choices but also on revenue of the rooms, travelling expenses, food etc. Then Girish et.al. [6] concluded that data mining is gaining popularity day by day for being applied in number of applications and one such application is travel and tourism. The data mining simply means to become more and more intelligent with the help of examples. By applying some well-known data mining techniques to tourism we can improve it up to a large extent. After 2015 yehong et.al. [7] from China did an analysis of tourism articles in S.S.C.I. journals from 2001 to 2012. He emphasized on the importance of journals for improving tourism sector, research counterparts and their work. Valeriya et.al. [8] use a new data mining technique first time termed as Decision Trees in their research for tourism. He used it to analyse the behaviour of inbound tourist to Japan and later on, use it for future destination marketing. This paper gives insights on how the decision trees are helpful in

providing the tourism sector a new strategy to improve it further. Another paper in which Krittipat et.al.depicts about the agro-tourism in their work. He renders how the agro-activities are helpful in improving the tourism of a Thailand. Some activities like reaping and tasting the fruit, walking, feeding animal in the orchard is some of the most liked activities. It aimed to stimulate the tourism by applying association rule. [9] Lastly, George et.al. [10] worked on predicting the tourism of Florida using data mining techniques. He used a forecasting technique M.A.D. (Median Absolute Deviation) to identify outliers in the system. It helped in predicting the future tourism of Florida. Therefore, these are the some insights about the research work done in the field of tourism using data mining. Apart from that, there is a lot of scope in the research area of data mining and it is beneficial to the tourism industry of any country.

VI. CONCLUSION AND FUTURE WORK

This paper has demonstrated about the different researches works in the field of tourism using data mining. The overall goal of the paper is to show the impact of data mining on the tourism sector. However, it is not the end of data mining in tourism rather it will have much wider applications in future. In our succeeding work, we will employ some new data mining techniques to predict the tourism and further improving the country's economy.

VII. REFERENCES

[1]. Rob Law, Henry Mok, and Carey Goh, "Data Mining in Tourism: A Retrospective Analysis," Springer-Verlag Berlin Heidelberg, ADMA 2007, LNAI 4632, pp.508-515.

[2]. Manoj B.Karathiya, Ravinder Singh Sakshi, Diler Singh Sakshi and Dhaval R.Karathiya, "Data Mining For Travels And Tourism," Research Gate, Journal of Information and Operations Management, ISSN: 0967-7762, Volume 3, Issue 1, January 15, 2012.Available: http://www.bioinfo.in/contents.php?id=55.

[3]. Pairaya Juwattanasamran, Sarawut Supattranuwong and Sukree Sinthupinyo, "Applying Data Mining to Analyze Travel Pattern in Searching Travel Destinations Choices," The International Journal Of Engineering And Science, ISSN(e): 2319-1813 ISSN(p): 2319-1805, Volume 2, Issue 4, pp.38-44, 2013.

[4]. Mirjana Pejic` Bach, Markus Schatten, Zrinka Marusic`, "Data Mining Applications in Tourism: A Keyword Analysis," Central European Conference on Information and Intelligent Systems, Faculty of Organizations and Informatics, pp.26-27, 2013.

[5]. Prabal Mahanta and Suhas Sudheendra, "The Impact of Data Mining for Insights in Travel and Tourism Industry," International Journal of Computer and Electrical Engineering, Volume 6, no.2, April 2014.

[6]. Yehong Sun, Yunjie Wei, Lingyun Zhang, "International Academic Impact of Chinese Tourism Research: A Review Based on the Analysis of SSCI tourism articles from 2001 to 2012," Elsevier, Tourism Management xxx, pp.1-8, 7 March 2016.

[7]. Girish Kumar Sharma, Promila Sharma, "A Study on Data Mining Algorithms for Tourism Industry," International Journal of Latest trends in Engineering and Technology, ISSN: 2278-621X, Volume 7, Issue 1, May 2016.

[8]. Valeriya Shapoval, Morgan C.Wang, Tadayuki Hara, and Hideo Shioya, "Data Mining in Tourism Data Analysis: Inbound Visitors to Japan," Journal of Travel Research, pp.1-14, 2017.

[9]. Krittipat Pitchayadejanant, Parinya Nakpathom, "Data Mining approach for arranging and clustering the agro-tourism activities in orchard," Elsevier, Kasetsart Journal of Social Sciences xxx, pp.1-7, 5 July 2017.

[10]. George Kurian, Hongmei Chi, "Predict Florida Tourism Trend via Using Data Mining Techniques," PEARC17, ACM ISBN 978-1-

4503-5272-7/17/07, New Orleans, LA, USA, July 09-13, 2017.

[11]. Ana Cristina.(April 2017).Online].Available: https://www.wttc.org/media-centre/press-releases/press-releases/2017/Indias-is-the-worlds-7th-largest-tourism-economy-in-terms-of-gdp-says-wttc/.

[12]. Saed Sayed.www.saedsayed.com.Online].Available: www.saedsayad.com/decision_tree.htm.

[13]. Faraz Bukhari www.tutorialspoint.com. Online].Available: https://www.tutorialspoint.com/data_mining/dm_knowledge_discovery.htm

[14]. www.wideskills.com Online].Available: www.wideskills.com/data-mining-tutorial/data-mining-architecture