

# Knowledge Discovery for Tourism Using Data Mining and Qualitative Analysis: A Case Study at Johor Bahru, Malaysia

*Atae Rezaei Aghdam, Department of Information Systems, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*

*Mostafa Kamalpour, Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*

*Alex Tze Hiang Sim, Department of Information System, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*

---

## ABSTRACT

*This paper aims to propose a new guideline for analyzing tourist profiles as found in [www.tripadvisor.com](http://www.tripadvisor.com). These have been examined from two different aspects so as to gain conclusive results. Tourist data were "crawled" from [tripadvisor.com](http://tripadvisor.com) through a specific web crawler. Mining techniques using a combination of visualization, clustering, and association rules were instrumental in discovering the first set of interesting knowledge. This was followed by a qualitative analysis applied through Nvivo software via coding of the tourist's comments in order to define the design of the prospective model. A final set of results was obtained once both results confirmed each other. In this study, results show that there are several types of tourists; with each group having different preferences. For example: male Singaporean visitors to hotels tend to enjoy wine and food in addition to outdoor activities; while local visitors to Legoland are not satisfied with certain aspects, such as the price of food. International tourists, however, consider the affirmative points of Legoland. This research can be very useful for tourist associations and hotel managers in Johor Bahru.*

*Keywords: Data Mining, Johor Bahru, Knowledge Discovery, Malaysia, Qualitative Analysis, Tourism*

---

## 1. INTRODUCTION

According to the annual report of Tourism Malaysia in 2010 (Tourism Malaysia, 2010), the tourism industry has played an important role in increasing the GDP with some 24.6 million

tourists arriving in Malaysia. In addition, the World Tourism Organization estimates that global tourism is expected to grow faster than other economic sectors in the world (Witten & Frank 2011). Uncovering new, interesting and useful information on tourist data can be

DOI: 10.4018/ijabim.2014100105

helpful for tourism organizations in order to identify tourists' behavior patterns and their preferences. This, then, is a great opportunity for the Malaysian tourism industry to boost tourist arrivals and increase revenue. This paper focuses on the role of data mining in the Malaysian tourism industry, in particular, regarding places of interest in Johor Bahru and hotels in the Mersing area. In essence, we will analyze tourist behavior patterns in Johor Bahru so as to discover useful and hidden knowledge in order to recommend appropriate places to visit. Moreover, quantitative and qualitative analyses have been undertaken simultaneously in this paper. Quantitative analysis was applied by Weka machine learning software; while qualitative analysis was performed by Nvivo software based on tourists' comments. Each group of analysis findings supports the other in order to obtain significant results. This study can assist tourist associations in Johor Bahru and travel agencies in the promotion of places in Johor Bahru which would be attractive to both local and international visitors. This paper is arranged in the following manner: Section 1; investigation of related works regarding data mining in the tourism industry; Section 2; explanation of our research methodology and pre-processing procedure; and finally, Section 3; analysis of the data and illustration of "hidden knowledge" in order to recommend these to tourists.

## 2. LITERATURE REVIEW

Currently, data mining supports various kinds of application tasks: from data pre-processing to association rules discovery, data classification, and cluster analysis respectively (Witten & Frank, 2011). Actually, it is part of the decision-making process; and the availability to analyze data automatically helps to determine a potential model. It also assists in estimating customer behavior in the realm of enhancing decision makers' ability to both adjust marketing strategy and reduce risks (Li, 2012; Han & Kamber 2006). In another study (Bose, 2009), concentration was given to

three main aspects of using data mining in the tourism industry. These are, namely; forecasting tourist expenditure, analyzing profiles of tourists and forecasting the number of tourist arrivals. The author has found various results based on these three dimensions. For instance, in forecasting tourist expenditure, artificial intelligence sources such as Neural Network were used for estimating tourist expenditure in the Balearic Islands. Further, Au & Law (2002) used data mining techniques to predict shopping expenditure by tourists with an accuracy level of 94%. In relation to analyzing tourist profiles (Bose, 2009), categorized tourists into specific groups such as; developmental support, prudent developers, ambivalent, cautious and protectionist respectively by using a clustering technique. In forecasting tourist arrivals, some studies had been conducted examining tourist arrivals to Hong Kong from six different countries; Artificial Neural Network (ANN) showed that this outweighed statistical methods. In a further study, Bose (2009) stated that it has been argued that, to date, only some AI techniques such as ANN and clustering techniques have been used in tourism data mining. It is largely prepackaged software that uses these techniques readily; it can also be used with little training to analyze data. However, the author believed that, in the future, more than one method will be applied for analyzing data. In this paper we focus on analyzing tourist profiles from two different perspectives, namely: quantitative and qualitative. In the following sections, the related studies illustrate how traditional data mining techniques have been used in the tourism industry.

### 2.1. General Research into the Tourism Industry

Data mining offers different approaches by which to gain knowledge and subsequently use this knowledge in the realm of, specifically: enhancing revenue, identifying customers' behavior patterns, recommendations to customers, etc. In the tourism industry in particular, a great deal of work has been devoted so far, but still

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

[www.igi-global.com/article/knowledge-discovery-for-tourism-using-data-mining-and-qualitative-analysis/126505?camid=4v1](http://www.igi-global.com/article/knowledge-discovery-for-tourism-using-data-mining-and-qualitative-analysis/126505?camid=4v1)

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Business, Administration, and Management, InfoSci-Digital Marketing, E-Business, and E-Services eJournal Collection, InfoSci-Select. Recommend this product to your librarian:

[www.igi-global.com/e-resources/library-recommendation/?id=2](http://www.igi-global.com/e-resources/library-recommendation/?id=2)

## Related Content

---

Intellectual Capital in Knowledge-Intensive Firms: Exploring the Concept and Main Components in Boston's Route 128

Pedro López Sáez, José Emilio Navas López and Gregorio Martín de Castro (2009). *Knowledge Ecology in Global Business: Managing Intellectual Capital* (pp. 224-235).

[www.igi-global.com/chapter/intellectual-capital-knowledge-intensive-firms/24924?camid=4v1a](http://www.igi-global.com/chapter/intellectual-capital-knowledge-intensive-firms/24924?camid=4v1a)

Women Entrepreneurship in Gulf Region: Challenges and Strategies In GCC

Viju Mathew (2019). *International Journal of Asian Business and Information Management* (pp. 94-108).

[www.igi-global.com/article/women-entrepreneurship-in-gulf-region/216922?camid=4v1a](http://www.igi-global.com/article/women-entrepreneurship-in-gulf-region/216922?camid=4v1a)

Collaborative Synergy and Leadership in E-Business

Kenneth David Strang (2011). *Global Business: Concepts, Methodologies, Tools and Applications* (pp. 1350-1374).

[www.igi-global.com/chapter/collaborative-synergy-leadership-business/54843?camid=4v1a](http://www.igi-global.com/chapter/collaborative-synergy-leadership-business/54843?camid=4v1a)

## Prioritizing Sectors for Economic Development in Sikkim, India

Manjushree Mishra and Ajeya Jha (2014). *International Journal of Asian Business and Information Management* (pp. 59-73).

[www.igi-global.com/article/prioritizing-sectors-for-economic-development-in-sikkim-india/114698?camid=4v1a](http://www.igi-global.com/article/prioritizing-sectors-for-economic-development-in-sikkim-india/114698?camid=4v1a)