หลักการและวิธีวิเคราะห์ข้อมูลทางธุรกิจในอุตสาหกรรมการท่องเที่ยว
และบริการด้วยวิธีการทำเหมืองข้อมูล

กฤษฏิพัทธ์ พิชญะเดชอนันต์*

มหาวิทยาลัยบูรพา ตำบลแสนสุข อำเภอเมือง ชลบุรี 20131

บทคัดย่อ

บทความฉบับนี้มีวัตถุประสงค์เพื่อให้เห็นความสำคัญของการวิเคราะห์ข้อมูลในยุคที่มีการใช้เทคโนโลยีเพิ่มขึ้นอย่าง รวดเร็ว ซึ่งข้อมูลถูกนำไปใช้อย่างกว้างขวางในวงการธุรกิจ โดยเฉพาะอย่างยิ่งในกลุ่มอุตสาหกรรมการท่องเที่ยวและการบริการ ในปัจจุบันมีการสร้างข้อมูลขึ้นอย่างต่อเนื่องและเป็นจำนวนมหาศาลโดยผ่านการใช้เทคโนโลยี อาทิเช่น โทรศัพท์มือถือ เว็บไซต์ สื่อสังคมออนไลน์ ซึ่งข้อมูลในด้านพฤติกรรม ทัศนคติ และกิจกรรมของคน ได้ถูกเผยแพร่และเก็บบันทึก ดังนั้น นักวิเคราะห์ข้อมูล ทางธุรกิจสามารถนำข้อมูลที่เก็บไว้มาวิเคราะห์เพื่อเข้าใจความต้องการและพฤติกรรมของลูกค้าเพื่อตอบสนองหรือนำเสนอ ผลิตภัณฑ์ได้ตรงตามความต้องการ อย่างไรก็ตาม ในกระบวนการวิเคราะห์ข้อมูล นักวิเคราะห์ข้อมูลจำเป็นต้องเข้าใจความ เหมาะสมของโครงสร้างของข้อมูล การสร้างรูปแบบพยากรณ์ การประเมินศักยภาพของรูปแบบ และการแปลผล บทความฉบับนี้ นำเสนอแนวโน้มของเทคโนโลยีที่เกิดขึ้นในกลุ่มอุตสาหกรรมการท่องเที่ยวและการบริการ แหล่งข้อมูล โครงสร้างข้อมูล ความ แตกต่างระหว่างวิธีการทางสถิติและการทำเหมืองข้อมูล การคัดเลือกรูปแบบการพยากรณ์ที่เหมาะสม ตัวอย่างการประยุกต์ใช้ วิธีการวิเคราะห์เหมืองข้อมูล และการประเมินศักยภาพของรูปแบบ ซึ่งจะเป็นประโยชน์กับนักวิเคราะห์ข้อมูลทางธุรกิจและ นักการตลาดในอุตสาหกรรมการท่องเที่ยวและการบริการเพื่อสร้างความได้เปรียบในการแข่งขันในยุคแห่งข้อมูลที่เป็นแบบดิจิทัล

คำสำคัญ : วิเคราะห์ข้อมูลทางธุรกิจ / การทำเหมืองข้อมูล / อุตสาหกรรมการท่องเที่ยวและการบริการ

* Corresponding Author : krittipat@scitech.au.edu, krittipat@go.buu.ac.th.

  อาจารย์ หลักสูตรการจัดการโลจิสติกส์ วิทยาลัยนานาชาติ

Emerging Business Analytics in Hospitality and Tourism Industry by using Data Mining Techniques

Krittipat Pitchayadejanant*

Burapha University, Saensook, Muang, Chonburi 20131

Abstract

This article aims to increase the awareness of the importance of data in the fast pace of technology. The usefulness of data is widely mentioned in businesses, especially in the hospitality and tourism industry. Technological devices are used in the industry and continuously create data through such technologies as smart phones, database systems, website and social media. The data on human habits, human activities and human attitudes are collected, shared and exchanged simultaneously throughout the society. It is indeed an opportunity for business analysts to gain more information in order to analyze the customers' behavior for understanding their insight, with the purpose of offering more attractive products and services to retain such customers. However, business analysts must understand the appropriateness of data structure, predictive model building, model performance evaluation and interpretation. This article describes the coming trend of technology emerging in the hospitality and tourism industry, sources of data, data structures, quality of data, differences between classical statistical modeling and data mining, how to select a suitable predictive model, application of data mining techniques and model performance evaluation. The article will allow business analysts and marketers to have a better idea on how to build a predictive model in the hospitality and tourism industry to maintain a competitive advantage in the age of digital data.

Keywords : Business Analytics / Data Mining / Hospitality and Tourism Industry

---

*   *Corresponding Author : krittipat@scitech.au.edu, krittipat@go.buu.ac.th.*

    Lecturer, Department of Logistics Management, International College.

Introduction

Thailand 4.0, proposed by Gen. Prayuth Chan-ocha, is a national policy to transform the Thai economy from "working hard to gain little" to "working a little but make big gains." With the policy, the Thai economy must be changed from industry-driven (Thailand version 3.0) to technology-driven (Thailand version 4.0). The effect of an innovative-driven Thai Economic structure is altered into a Value-Based Economy. The characteristic of a Value-Based Economy is producing the products and services integrated with technology, creativity, research and development, innovation and science and technology in order to enhance the value of products and services.

Currently, the concept of Thailand 4.0 has been implemented and fostered by both public and private sectors. Thailand version 4.0 policy emphasizes new developments in five industries: 1) Food, Agriculture and Bio-tech, 2) Health, Wellness and Bio-Med, 3) Smart Devices, Robotics and Mechatronics, 4) Digital, Internet of Things (IoT) and Embedded Technology, and 5) Creative, Culture and High Value Services. Hospitality and Tourism industry is in the fifth category of new emerging industries.

With the fast pace of the technology era, information technology has changed the way of travel in Thailand. People are browsing the Internet for travel inspiration and other travelers' photos, check-ins, ratings and more [1]. Nowadays, in the tourism industry in the ASEAN region, many visitors get to know about interesting travel destinations through the Internet and tourists' recommendations of what they have enjoyed through social networks. The prime minister hopes that e-travel industry will be of greater reliability and safety for tourists who use ICT to search for tour packages, book hotels, and discover new destinations. New technologies are adopted for supporting the hospitality and tourism industry to enhance the value and quality of services. Many data are recorded and analyzed for managing the relationship between businesses and customers, called customer relationship management (CRM). According to the trend, this management is essential in the industry, especially online booking websites such as Agoda, Booking.com, etc.

Customer relationship management is a competitive advantage for business. The business owners would like to build the relationship with customers in order to retain customer loyalty rather than having only high circulation of new customers. To explore what customers need, data are essential. Data will help to determine what kind of travelers visit your place, data can help business analysts tailor the marketing campaign to a specific target group, and data can help to understand exactly where your visitors come from. Tourists have an overwhelming demand to satisfy their unlimited desires. The business owners must understand and respond to their needs to attain the customers' loyalty to prosper in their business. The demands of tourists can be analyzed through the processes in a business's systems [2]: information inquiry system, online booking and payment system, consumer reviews and recommendation system, online consultation system, and special service management system.

For instance, the travelers booked the flight and accommodations through Online Travel Agents (OTA) rather than using a travel agency [3]. The data have been driven into Air Asia low cost carrier to investigate the customers' expectations [4]. Moreover, there are many travel bloggers to recommend interesting destinations. The comments or interactions through technology can be traced to discover the customers' behavior.

As a consequence, the organization needs the person who can evaluate and predict what customers

need based on the data obtained from many sources. A business analyst is the person who should understand and be able to construct the business model in order to achieve the business objectives. The business analyst should understand : the objectives of the business; the sources of data; how to get the data with high quality; characteristics of the data; what metrics should be used to analyze; the objective of each predictive model; the toolkits in predictive model analysis; how to evaluate the model's performance; and how to interpret the results.

Technological support in hospitality and tourism industries

Technologies that are transforming the hospitality and tourism industry consist of 1) Online booking systems [5, 6], 2) Electronic point of sale (EPOS) systems [7], 3) Customer relationship management (CRM) software [8], 4) Marketing automation [8], 5) Social media [9], 6) Smartphones [9], and 7) Smart appliances [9].

Online booking systems are widely used to reserve the room in hotel, the seat in the restaurant or purchase the ticket for a particular purpose before traveling. Presently, the third party online booking providers provide services to customers; they are convenient to reserve or purchase services remotely through their system via website or application. From the business providers' perspective, the employee can manage the plan in order to provide services in advance, such as procuring the raw material or preparing the merchandise to serve their customers' request on time.

Electronic point of sale system is sophisticated technology which is widely adopted to encourage ease of use and convenience in hospitality operations. For illustration, the service provider can monitor and control the vacancy of the room or seat abruptly; the

transactions of customers are simultaneously viewed in real time. In addition, the useful information of the business performance, in short and long time, can be reported to the stakeholders for assisting for their decision-making.

Customer relationship management (CRM) software is the tool to organize the customers' contact for creating a relationship with the customers, especially in marketing and sales departments. The vast majority of customers' information and their behavioral information are recorded and organized systematically into the system, ready to be retrieved as being as the recognition system. In accordance with CRM software, the customers are treated differently depending on their behavior and lifestyle. It is necessary for marketers to support their decisions based on consumer needs information. Therefore, it not only satisfies consumers but also ensures that they will become loyal customers [10]. The system can automatically operate to support employees to manage their customers with the recorded data. From the stored data in CRM, business analysts can extract the diamond information, which can assist the business to understand the hidden information that might support its success.

The indicators of hospitality and tourism to manage the customer relationship have been grouped into three categories [11]: economic performance indicators, customer behavior indicators, and customer perception and experience indicators. The economic performance indicators are the indicators of hotel and facilities performance, comprised of booking information, overnights, prices, occupancy, sales, etc. Customer behavior indicators emphasize behavior in website navigation and browsing, booking and consumption of services, and customer profile. Finally, customer perception and experience indicators focus on the evaluation of customers after they receive the services

comprised of perceived destination brand awareness, destination values, value for money, customer satisfaction, and loyalty [12].

Marketing automation is the system that allows the emails to be sent to customers or prospects personally, which are classified based on their interest and behavior. The information appended into email can be promotions such as redeemed coupons, new offered services, or news and events. The data are essential to evaluate the customers and prospects for making the personal interest classification in order to use a marketing automation system. For instance, booking.com or tripadvisor.com, etc. send emails to their customers about the promotion of packages to encourage them to purchase the packages through their websites.

Social media is currently adopted in many businesses to discover the opinions or attitudes of customers as the feedback of their business. Nowadays, social media has high influence and impact on the change of consumer behavior. Data are created a lot in social media and are beneficial for business analysts to extract the diamond information. Data are created tremendously within a second from all social media users [13]; over four million posts every minute on Facebook, about 1,736,111 likes on photos each minute of the day in Instagram, and each minute, Twitter users generate 347,222 tweets, and three billion video views per day on YouTube. In Thailand, social media users on Facebook, Instagram and Twitter are rising to 47 million, 11 million, and 9 million, respectively [14]. In the hospitality and tourism industry, text messages, images and videos posted on social media can be useful information that can be analyzed by business analysts. Many travelers post nice images and share their positive and negative opinions and attitudes related to the particular destination on social media. As a consequence,

business analysts can analyze the characteristics of tourist destinations where travelers would like to travel from social media. In addition, blogs are another source where business analysts can gain information about attitudes toward a particular travel destination. Everybody can create a blog for sharing either positive or negative experiences in travelling or asking questions for resolving doubts such as directions to a destination, or how to solve problems when facing a situation abroad, and reviewing the particular travel destination.

Smartphones are the technological devices which support the hospitality and tourism industries. It has high effects toward the mobility of social media because travelers mostly share their images, post comments, and express opinions or feelings when they arrive at the destinations through smartphones.

Presently, social media applications are default installed into smartphones, such as Facebook, Line, etc. As a consequence, the travelers can access social media comfortably. Because of sharing the comments and images, the data are generated through social media by travelers, which business analyst will be able to analyze for sentiment or content.

Smart appliances or Internet of Things (IoT) are the electronic appliances that are embedded with cloud or Internet connectivity systems which can be controlled by a single device such as a tablet or smartphone. In a hotel, there are many appliances that can be controlled by smartphones such as light, air conditioning, TV, room service, etc. Even in sightseeing, digital devices are also adopted to increase the impression of travelers' experience. The smart appliances can be used to record the data of guest preferences based on their activities.

For instance, Mandarin Oriental Hotel in New York City can track all guests' preferences through the smart appliances or Internet of things such as the light, radio, and refrigerator because data have been

recorded through these appliances when they use them. The appliances do so automatically based on their recorded data. For instance, the electrical light dims automatically for being comfortable to customers; the preferred music turned on when the guest checks in; and sensors in refrigerators alert the maids at the minibar. In addition, Starwood Hotel enhances the guest experience by installing smart mirrors. It is a touchscreen from which guests can check email, sport scores, weather, and Twitter feed on the mirror. As a consequence, the behavioral data are recorded based on smart mirrors.

Due to technological developments, business analysts can determine the characteristics of travelers and extract the utilized information through the data obtained by technology. However, even when the data are provided, the key is how to select the correct data and analyze them to explore and understand the customers' needs. With the various technologies, data are not solely recorded in a single pattern which will be mentioned. Nowadays, the data structure is complicated to analyze.In Thailand, nowadays, the infrastructure of collecting data in the hospitality and tourism industry is not united and completed, especially in the government sector. The infrastructure of ICT is weak at the moment [15]. The business analysts need to purchase the data from the private agencies or collect data by themselves through databases, social media or websites such as Facebook, Agoda, Airasia, etc. However, the quality of data is widely mentioned and imperative for predictive model building.

With the various technologies which create the massive amount of data, presently, "Big Data" has been mentioned widely, especially in the business sector. Big data focuses on the enormous size of data in storage. However, it is not just about the volume; it is also the velocity and variety of data.

With velocity, the data have been updated and recorded simultaneously – real time. For variety, the structure of data is not identical; it has been codified into three taxonomies: structured, semi-structured, and unstructured data.

Data Quality

Input data with good quality will cause the high quality of the outcome information; garbage in, garbage out. The business analyst must identify how to obtain good quality data. To assess the quality of data, there are six dimensions [16]: completeness, uniqueness, timeliness, validity, accuracy, and consistency. For completeness, the data should be 100% completed – without any spaces, blanks, or missing any observations. Sometimes, imputation from the available data is applied to fulfill the missing values. Or else, the observation which has missing values will be discarded. Secondly, one observation should be recorded once, not repeatedly collected in a particular study - uniqueness. The observation should not be duplicated; it will cause bias and the obtained information will not be generalized or influenced by repeated observations. Thirdly, timeliness: data should be recorded at the suitable time based on the objective of study to prevent outdated data. Fourthly, validity is mentioned regarding the format of data. It is imperative to identify whether they are structured, unstructured or semi-structured data. Unless, the format of data is defined, data management before analysis will be complicated. Fifthly, accuracy: the data are really being understood by the observation. Before collecting data, each metric should reflect the real world object and be understandable. For instance, the format of DD/MM/YYYY or MM/DD/YYYY should be clarified. Otherwise, there will be faults and pitfalls in interpretation.

Lastly, consistency: the data should be consistent when they are collected from various locations. Otherwise, business analysts cannot compare the results from different locations. As for the dimension of data quality, data structure is one of the dimensions to be identified as the quality of data.

Data Structure

To be ready for Thailand version 4.0, various technologies create different patterns of data. Business analysts have to understand the pattern of obtained data; otherwise, the analyzing techniques for knowledge discovering will be misused. Data are classified into three patterns: structured, unstructured and semi-structured [17].

Structured data can be reformatted and organized into fixed fields within a record. It is split into small units. The data can be contained in tables, which are easy to access via relational databases or spreadsheets. It has been recorded into specific rows and columns. It is simply searched and reached. Usually, the data are recorded as the attributes into databases. The database developers and team plan and create what data should be kept for being of benefit for further analysis. The drawback of structured data is it represents only 5 to 10% of all informatics data [18].

Unstructured data refers to the unorganized data; not in the fixed field; and unable to be recorded into the spreadsheets. It is the opposite of structured data. Typically, unstructured data can be text, audio, or image file format. Lots of social media usually record this type of data. For illustration, social media users comment their opinions, feelings, and attitudes and also show their images through Facebook, blogs and also review the accommodation into websites. Because of technological growth, unstructured data have prospered unexpectedly; about 80% of data is unstructured [19].

Semi-structured data are between unstructured and structured data. For instance, Text: XML or electronic data interchange messages contain tags and are able to identify the category of the topic in each day. Secondly, web server logs: the business analyst can journey through a web server log in order to search the interesting content to find out the trends of consumers in detail. The amount of this type of data is also represented in the few parts of data. With the incoming of different structures of data, data source is determined by the incoming data. Table 1 shows the example of data sources which new technology adopted in the hospitality and tourism industry in order to collect data. Different data sources provide different data characteristics. The analyst must understand what data can be transformed, what appropriate attribute should be selected and how to analyze for benefit.

Table 1 Example of data in hospitality and tourism industries [14]

Data Source	Technological Devices	What data is provided?	Level of Structure
Customer transactions and account records	- Online reservation through the smart devices such as smartphones - Customer relationship management software	- Geo-demographics information - Contact details - Purchase history - Payment history	High structure

Data Source	Technological Devices	What data is provided?	Level of Structure
	- Electronic point of sale		
Customer relationship management systems	- Database system - Online booking system - Website - Social Media	- Inbound and outbound of customers' and hotels' information	High structure
Web-logs / Clickstreams / Page tagging	- Website - Social media	- Path of customers' searching to a site (tracking cookies) - Pages, concerned hospitality and tourism industries, visited within a site - Time duration spent in each page - IP address of computer used to search which can detect geographical location	High structure
Power dialers	- Smartphone	- Date, time and length of phone conversations - Recordings of calls and outcomes	Medium structure
Sensors	- Smart appliance such as Air conditioner - Google Map - Smartphone	- GPS - Preferred Room Temperature	High structure
Mailing lists	- Customer relationship management software	- Contact details of customers	High structure
Customer correspondence	- Customer relationship management software - Marketing automation	- Message in e-mail about what the customers' request	Low structure
Cameras and imaging systems	- Social media data such as YouTube, Facebook - Website - Online booking systems - Video and surveillance camera	- Pictures - Videos	Low structure

Data Source	Technological Devices	What data is provided?	Level of Structure
Surveys	- Smartphone - Website	- Mixture of geo-demographic, behavioral, attitude, and sentiment data	Depends on survey design.
Websites / pages	- Website	- Text, tables, pictures, and video	Low structure
Social networks	- Social media	- Sentiment data - Emoticon - Geo-demographic	Low structure
Tweets and blogs	- Social media such as Twitter - Website such as Pantip.com - Blogs	- Text what customer want, how they feel, what they thought	Low structure

High structure of data represents the well-structured data for being ready to do predictive models. However, data cleaning is required. Low structure data require many processes to transform into structured data in order to be ready for constructing predictive models. With the vast amount and various data structures, traditional statistical modeling cannot support business analysts in analyzing the data and obtaining useful information. Nowadays, data mining techniques are proposed widely to support the business analysts discovering the knowledge obtained from customer behavior. In addition, data mining techniques can be applied to construct the inherent recommendation system in tourism such as Personalized Sightseeing Planning System (PSiS) [20, 21], which can be intermediary between customer and travel agency [22], to provide information for a specific customer.

Classical Statistical Modeling versus Data Mining

To demystify the differences between statistical modeling and data mining, there are four aspects of the differences [23]: *Firstly*, the classical statistical modeling focuses on hypothesis testing theory by using samples, whereas data mining techniques use machine-driven model building and the data are gathered with the big size as being close as the population. *Secondly*, data mining techniques are applied for building models in order to predict, while classical statistical tools examine the preconceived notions of researchers. *Thirdly*, data mining produces enormous gains in terms of performance, speed of use, and user friendliness; however, data miners must understand principles of statistics in order to build the model. *Fourthly*, the outcome based on traditional statistical analysis is usually biased and misleading when the business analysts violate the assumptions. It affects the results. However, data mining techniques focus on the performance of model prediction.

Application of Data Mining Techniques in the Hospitality and Tourism Industry

The data in previous research studies are retrieved from the emerged technologies: database, comments through social media, travel websites, blogs, and customer relationship management systems. Currently, these technologies provide tons of data which are waiting to be analyzed with data mining techniques. Various data mining techniques that are widely applied consist of association rule, regression analysis, classification tree, clustering analysis, text analytics, and multi-analysis approach.

Association Rule

Association rule is the technique to discover the sequence or relationship among items that will be selected over time. It estimates what product items will be purchased together. For instance, emerging hotel preference patterns are investigated [24]. The aggregated data are online reviews based on hotels' concerns which were posted on the TripAdvisor website from 2010 - 2013. The patterns of hotel preferences in five countries are investigated: Hong Kong, Singapore, Shanghai, Bangkok, and Sydney. The English comments concerning hotel facilities are tokenized and transformed from unstructured data into structured data by coding the value 1 as being mentioned in the review and 0 otherwise. Thereafter, Emerging Pattern Mining (EPM) technique detects the changes in customer segments and discovers user behavior patterns of hotel features.

In Thailand, the travel pattern is investigated at Chiang Mai. The travel pattern in searching consists of three main categories of travel destination choices: accommodation, tourist attractions, and things to do are analyzed by using the association rule technique [25]. The result of the study is applied for developing mobile applications to generate a pattern of recom-mended destination choices for travelers.

In agro-tourism, the choice of interested activities are listed in order to find out the association among them [26]. Frequency pattern algorithm is applied and evaluated by support, confidence and lift values. The result of the study can recommend the pattern of interesting agro-tourism activities to the tourists.

Electronic word of mouth (eWOM) is the electronic feedback communication system which can be adopted to improve service quality and marketing strategies. An existing study discovered the user groups' identification by using two types of attributes: demographic attributes such as age, gender, education, income and travel experience, and travel information attributes such as past travel destination, travel motivation and online experience [27]. With these characteristics, the researchers can create the patterns and characteristics of travelers' information browsing and sharing based on the association rule into four taxonomies: 1) browse and share, 2) browse and not share, 3) share and not browse, and 4) neither browse nor share. The advantage of the outcome to business analysts is the ability to understand the characteristics of customer, feedback of sharing their opinions and creating the contents in order to improve service quality.

Regression Analysis

Regression analysis is widely applied in hospitality and tourism. It is the technique to find out the model of relationship among variables and predict the outcome. Multiple linear regression is used when the characteristics of a set of input variables (predictors) and a single output variable (response variable) are continuous variables or numeric variables. For illustration, the overall satisfaction is estimated by a 5-point Likert scale on five aspects: hotel location, cleanliness, room experience, service quality, and

value for money [28]. The severe problem of regression, multicollinearity problem, has been analyzed with the threshold of variance inflation factors (VIFs) of 5 before regression analysis [29]. They found that all five dimensions are significant to predict overall satisfaction.

Logistics regression is a technique to study the effect between a set of input variables which can be either numerical or categorical data and a single binary output variable. This technique is applied to predict Hong Kong online browsing or purchasing habits by using traveler perception factors, demographic factors, and traveler planning factors as predictors [30]. The researchers use the demographic information: education, age, experience, income, and tourists' behavior: browser and nights on trip, to identify the segment and compare the "Actual Target %" and "Predicted Prob %" whether there are significant differences between two independent samples. If there is no significant difference between "Actual Target %" and "Predicted Prob %", the model developed can provide good estimation. The advantage is the online marketer can approach the internet customer for offering the tailored service.

Classification Tree

Classification is the technique to classify the customer based on the data of each observation in dependent variables that have been recorded into category; data are supervised. The target categorical variable is used as the classification model construction in order to predetermine classes or categories. There are many popular classification trees such as Chi-square Automatic Interaction Detector (CHAID) algorithm, C4.5, C5.0, and Classification and Regression Trees (CART). The advantage of classification tree is easily understandable. The rules are presented through the tree. Furthermore, the business analyst can identify or rank the importance of classification variables based

on this technique.

The existing study applied the classification tree method – Chi-square Automatic Interaction Detector (CHAID) algorithm, to explore tourist loyalty intention in the hotel sector [31]. The important variables to classify customers toward becoming loyal include physical environment and social interactions which are evaluated with five-point Likert scale. Meanwhile, tourist satisfaction level and loyalty intention toward hotel service, the demographic characteristics and travel-related information are measured as categorical variables. Another study using the classification tree in fast food restaurants to predict customer satisfaction in fast food restaurants [32] consists of ten input variables: taste, serving, dessert, restaurant environment, washing basin, waiting time, staff behavior, menu pricing, discount, and store location. According to the results, the essential variables that affect customer satisfaction are staff behavior, pricing, and restaurant environment.

Furthermore, customer churn behavior prediction is studied based on the decision tree method with Exhaustive CHAID (E-CHAID) technique [33]. The customer behavior variables: frequency, monetary value, length of relationship, interplay and bonuses/ rewards, and demographic variables: age, sex, and location, are input variables to construct the predictive model of customer churn. The accuracy of the model is evaluated by a classification accuracy table or confusion matrix table which will be mentioned in the model performance evaluation section. The advantage of studying customer churn is the manager can predict and prevent the customer leaving to purchase the services of other companies.

Clustering Analysis

Clustering analysis is the technique to classify the observations into categories such as classifying customers who have similar characteristics into the

same category. Cluster analysis differs from classification technique; cluster analysis will be used when the data are unsupervised data. There is no pre-classified category. The business analyst is able to classify the customer when the response of each observation is unknown by applying clustering analysis. Cluster analysis can classify the customer when the characteristics of the customer are known and used, which is called doing segmentation. For instance, this technique can segment the characteristics of a customer who prefers to look for information in Las Vegas and search for information on a particular restaurant through web log [34].

The algorithms in clustering are divided into seven families [35]; however, the popular methods are hierarchical agglomerative and iterative partitioning. In the hierarchical agglomerative technique, the concept is merging individuals by starting with each respondent represented in one group and ending with one single large group. The number of clusters is not determined in advance, unlike the iterative partitioning method. The iterative partitioning method starts with a random splitting of the observations and reallocates the observations based on pre-defined criterion optimization which is considered by minimum variance within the cluster. With this method, the number of clusters must be identified. K-means clustering is a common technique in the iterative partitioning method and is adopted after the business analyst obtains the suitable number of clusters.

The number of clusters to be defined has no satisfactory solution. Nowadays, the methods to identify the number of clusters are 1) heuristic procedure, 2) combination between heuristic and subjective opinion and 3) purely subjective manner. However, no matter what technique is selected, the result of defined clusters must be explainable.

Data visualization is another technique that assists the business analyst in finding out the suitable number of clusters when the number of clusters is not able to be identified. The mentioned data visualization technique is called Self Organizing Maps (SOM). Dursun and Caber [31] used SOM to profile customers in the CRM system of hotel chains. They can divide the number of clusters into eight clusters by RFM as being as the indicators for clustering [36]. RFM analysis indicators consist of Recency (R) value, Frequency (F) value, and Monetary (M) value. The information in each cluster is explored and discovered by K-means algorithm. They classify the profitable customer into eight taxonomies: 1) Loyal customer, 2) Loyal Summer season customer, 3) Collective buying customer, 4) Winter season customer, 5) Lost customer, 6) High potential customer, 7) New customer, and 8) Winter season high potential customer. For the advantage of study, the business analyst can lead the hotel to generate new strategies for CRM performance augmentation in each group to retain the customer.

Text Analytics

With the variety of social media, the comments and posts on social media are unstructured data. Text mining is beneficial knowledge for hospitality operation. In the past, unstructured data are costly to summarize and interpret by hand [37]. Nowadays, sentiment analysis is the technique applied for discovering and recognizing positive and negative expressions of human opinion. The purpose of this technique is to know the potential drift in society through texts or messages posted in social media. Business analysts are able to study the attitudes, observations, and expectations of stakeholders [38].

For instance, Latent dirichlet allocation (LDA) is a method being proposed for constructing topic modeling [28]. The method enables the discovery of topics from unstructured text data from online reviews

to find out how the customer rates the hotel. It is utilized to extract the dimensions of customer satisfaction, the importance of different dimensions, and the words related to the dimensions based on preprocessed comments in TripAdvisor.com [28]. Customer satisfaction in hospitality venues is analyzed by content analysis [39]. Key factors of customer satisfaction that are mentioned the most are listed and compared between luxury and budget hotels. The mentioned key factors consist of logistics, facilities, reception services, food and beverage management, cleanliness and maintenance, and value for money.

Schuckert [35] proposed two approaches to opinion mining: valence and feature extraction. Valence is categorizing online reviews into positive and negative aspects. Feature extraction is mining the word patterns about products and service by using keywords, topics, and concerns of customers. Capriello [36] demonstrated mining consumers through online travel reviews with sentiment analysis. They compared the farm tourism experience among four countries: Australia, Italy, United Kingdom, and United States of America. The willingness of reviewers to generate reviews of hotels online is assessed [40]. The person who reviews the products or services is imperative; a customer-generated review has positive association with popularity which totally contradicts with an editor-generated review [41].

Social media like Twitter as the data collecting tool is a powerful source for business analysts for extracting the insight information from tourists' behavior. The messages posted and disseminated into Twitter are explored for finding out the perception of Asian restaurants [42]. They select the useful English keywords and classify them into positive and negative evaluations to evaluate and compare the positive and negative comments among Asian restaurants. The five search words evaluated are Asian restaurant, Chinese restaurant, Japanese restaurant, Korean restaurant and Thai restaurant.

Booking.com is a large website to book the hotel. More than 345,000 hotels are listed for tourists' selection. During years 2014 to 2015, 19,318 hotels were reviewed through this website. The text mining approach is applied to extract reviews about hotels [43]. The room issues are reported as primary, such as room appearance, hotel appearance, cleanliness, interior design, and etc. Second are quality of staff and service, such as staff performance and personalization. Third would be added value such as free Wi-Fi, breakfast, and restaurant. Finally, the close proximity of hotels to attractions is considered.

Multi – Analysis Approach

Multi-analysis approach is applied to gain more information from the customer. Using an individual technique is not enough to gain more deep information. The researcher uses more than one technique to explore and discover customer insight. For instance, Liao et al. (2010) found out the knowledge pattern of customers for developing new products in tourism by using the Apriori algorithm [44]. The Apriori algorithm is one of the techniques to discover the association of one item to another – creating association rule – in data mining. After the Apriori algorithm is obtained, they applied cluster analysis to segment the Taiwanese tourists into five clusters to develop the market strategy regarding customer value and satisfaction. As a consequence, the business analysts are able to create a new business model for attracting new tourists.

In addition, a multi-analysis approach is also applied to discover the knowledge with both quantitative and qualitative methods. For illustration, regression and text analysis approaches are adopted [45] for studying the performance of hotel operation and eWOM. Simple

regression analysis is applied to analyze the hotel operating performance. The relationship between hotel operating performance and Internet exposure rate is examined. The hotel operating performance is measured by hotel operation income. The result of regression analysis shows there is a significant relationship between the Internet exposure rate and income of the five-star hotel, whereas text analysis shows the supportive reasons to motivate the reviewers' impression of becoming a customer in a particular hotel.

According to the reviewed studies, the data are collected via social media, websites, web-logs, and customer relationship management systems. These technologies are accepted as having a tremendous pace and the obtained data occur enormously and concurrently. However, there are many algorithms and techniques in data mining for building predictive models. The question the business analyst must ask is which technique should be selected after predictive models are built. To select the suitable predictive model, the business analyst has to be aware of predictive model performance evaluation.

Model Performance Evaluation

To evaluate the performance of the model, there are three techniques of evaluation classified by the objective of the modeling technique and characteristics of predictive variables. The three techniques consist of evaluating association rules, evaluating counting prediction, and evaluating numeric prediction.

1. Evaluating Association Rules

The goal of selecting rule is selecting the value that shows strong dependence between antecedent and sequent item sets. Normally, based on Association rules, there are three measurements to select the rule: support, confidence, and lift. The business analysts have to combine all three criteria for analyzing together in order to obtain the best rule.

Support is counting frequent one-item sets: how many transactions in the database with the item. It presents the percentage of the total number of records in the database. It contains all items in the antecedent and the sequent. With high value of support, the item set is selected more often. Hence, it means the set has high frequency of selection [46].

$$Support = P(A \cap B)$$

Confidence is a measure of the accuracy or reliability for a given rule. The higher value presents the more likelihood of the second situation occurring based on the first situation. Actually, the value of confidence estimates the conditional probability of the second situation given the first situation.

$$Confidence = P(B|A) = \frac{P(A \cap B)}{P(A)}$$

These two values should be greater than the minimum criteria which have been set up by business analysts or expert judgment. There is no rule of thumb for the suitable minimum value. Confidence level of more than 0.7 has been studied [26].

Lift is the measurement of the relative of selected one item with another item. This value reduces the possible bias incurred when using above standard measurements [44]. The interesting value of lift should be different from 1. The value that is close to 1 indicates event A and event B are independent events.

$$Lift = \frac{P(A \cap B)}{P(A)P(B)}$$

This evaluation technique is applied when the business analysts use the Association Rule in data mining technique to discover the relationship among the items or list of interesting activities in a tourism destination.

2. Evaluating Counting Prediction

Evaluating counting prediction is the evaluation technique when the response variable is a categorical variable after the observations are predicted and classified into category. The confusion matrix is constructed to compare the correctness and wrongness of predictions with real observations. The data mining techniques that should be evaluated by counting prediction are classification tree and clustering.

Confusion Matrix is the matrix that presents the classification results: frequency of correct and incorrect predictions [47]. The actual values from data sets are compared with the predicted values in the trained model.

Table 2 Confusion matrix table

		Predicted Class	
		Yes	No
Actual Class	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

According to Table 2, True Positives (TP) and True Negatives (TN) are correct classification counts while False Positive and False Negative are incorrect predictions. False Positive (FP) is the counting frequency when the outcome is predicted as yes (or positive) whereas it is actually no (negative). False Negative (FN) is the counting frequency when the outcome is predicted as negative while it is actually positive. From confusion matrix, the measurement values have been calculated:

$$Sensitivity\ (Recall) = \frac{TP}{TP+FN}$$

Its value should be high.

$$Specificity = \frac{TN}{FP+TN}$$

Its value should be high.

$$Positive\ Precision = \frac{FP}{TP+FP}$$

Its value should be low.

$$Negative\ Precision = \frac{FN}{TN+FN}$$

Its value should be low.

$$Accuracy = \frac{TP+FN}{TP+FP+TN+FN}$$

Its value should be high.

Cross Validation Error Rate (CVE rate)

$$= \frac{FP+FN}{TP+FP+TN+FN}$$

Its value should be low.

$$Overall\ success\ rate = \frac{TP+TN}{TP+FP+TN+FN}$$

Its value should be high.

$$Error\ rate = 1 - \frac{TP+TN}{TP+FP+TN+FN}$$

Its value should be low.

The performance indicators are also applied to compare in order to select the suitable algorithm [48]. The suggested indicators consist of computing time, sensitivity, specificity, accuracy, precision, and CVE rate. The procedures for selecting algorithms are:

1. Filtering by low computing time, then
2. Filtering by low precision value (approximately < 0.8), then
3. Filtering by CVE rate (approximately < 0.18), then
4. Filtering by highest accuracy, or
5. Filtering by highest overall success rate or lowest error rate [47]

The data mining techniques that the model should be evaluated by using confusion matrix are logistic regression or classification tree because the response or dependent variable of these techniques is categorical data.

3. Evaluating Numeric Prediction

Evaluating numeric prediction is the evaluation technique when the response variable is a numerical variable. The error in each observation can be calculated by the difference between predicted value and actual value. The model that causes the lowest error will be accepted. The example of data mining technique that should be evaluated by using numeric prediction technique is regression analysis.

All error measurement measures the difference between predicted value and actual value. In the model with good performance, the predicted value should be as close as the actual value. With the least difference, the error measurement value should be close to 0. When business analysts would like to use various algorithms to construct the predictive model, the model that produces the least error measurement would be selected.

The measurements of error consist of mean-squared error (MSE), root mean-squared error (RMSE), mean absolute error (MAE), relative squared error (RSE), root relative squared error (RRSE) and relative absolute error (RAE).

Correlation coefficient (R)

In addition, correlation coefficient can be applied to evaluate the performance of the model. The value shows the relationship between prediction values and actual values. The value of correlation coefficient shows the relationship between two variables. Its value lies between -1 and 1 whereby the value that is close to 1 indicates positive relationship; close to -1 indicates negative relationship; and close to 0 indicates no relationship.

With good performance of the model, both values should be similar. Therefore, the value should be close to 1 as the good performance model. Hence, the negative value should not occur in such a prediction method.

The data mining techniques that the model should be evaluated by using measure of error or correlation coefficient should have the response variable or dependent variable as quantitative data. The techniques that have quantitative data as the response variable are multiple regression, support vector regression, etc. In a previous study, Cankurt and Subasi [46] compare the performance of a model to forecast tourism demand through three models: multiple linear regression (MLR), multilayer perceptron regression (MLP), and support vector regression (SVR) [49]. Three

indicators to measure the model performance are introduced which consist of relative absolute error (RAE), root relative squared error (RRSE) and correlation coefficient (R). They select the model which has the highest R and lowest RAE and RRSE. According to the result, SVR shows the best performance.

However, rather than performance evaluation, explicability and simplicity are also important to be considered. The model should be explainable - why it happens and how it happens. Some data mining techniques are black box methods which are difficult or unable to explain the cause and effect between independent and dependent variables. This is the reason association rule, regression analysis and classification tree are popular in analysis because the business analysts are able to explain and clarify the reason of occurrence easily.

Furthermore, the performance of the predictive model should be stable and not deteriorate over time. With stability of the constructed model, the business can save the cost and time to rebuild the model and evaluate the performance of model. Nevertheless, the predictive model must be changed over time depending on ability of prediction. Business analysts should select the stable models such as simple linear model and decision trees [50].

Lastly and importantly, the predictive model is worthless when the business objectives do not accept. Especially in business, the predictive model must be understandable, explainable and rely on the business objectives. Each input variable is measurable, acceptable and definable. The biggest mistake is considering the predictive accuracy rather than applicability in business. As a consequence, business analysts must recognize what reasons there are to build the predictive model and for whom.

Conclusion

This academic article describes the revolution of data sources, technology support and data structure in the hospitality and tourism industry. With the revolution, the advanced data analytic technique which is called data mining is introduced, especially in business.

However, the hospitality and tourism industry incubates many businesses related with the industry. They have tremendous data obtained by customers. The applications of data mining techniques that are widely applied in the industry are association rule, regression analysis, classification tree, cluster analysis, text analytics, and multi-analysis approach.

As indicated in this article, data mining techniques and performance evaluation techniques are appropriate for the structured data. With current unstructured and semi-structured data analytics, they show only the descriptive information. They are not adopted to construct the predictive model yet. The predictive models of structured data are association rule, regression analysis, classification tree, and cluster analysis. Hence, their model performance can be evaluated. With text analytics, the data are unstructured; hence, the business analytics can present customer insight by showing descriptive information.

Lastly, the model performance evaluation techniques are introduced in order to assist the business analysts in making decisions in model selection. The technique of evaluation is selected based on the objective of study and characteristics of response variables or dependent variables. Also, the criteria to select the suitable model are listed and recommended in this article.

References

1. Higgins, A., 2016, 5 Ways Social Media Has Transformed Tourism Marketing [Online], Available : https : //hospitalitytech. com/5-ways-social-media-has-transformed-tourism-marketing [12 June 2017].

2. Zhang, Y.B., Wang, D. and Wu, J., 2014, "Research of Data Mining Technology under Tourism Information," *Applied Mechanics and Materials*, 687, pp. 1260 - 1209.

3. Olenski, S., 2015, Are OTAs Really Killing Brick And Mortar Travel Agencies? [Online], Available : https://www.forbes.com/ sites/steveolenski/2015/04/27/are-otas-really-killing-brick-and-mortar-travel-agencies# 53fb3fbb11d7 [20 July 2017].

4. Company, G.E., 2017, "Improve Operational Efficiencies and Customer Expectations [Online], Available : https://www.ge.com/digital/stories/data-and-analytics-driving-success-at-air-asia [25 July 2017].

5. Kucukusta, D., Law, R., Besbes, A. and Legohérel, P., 2015, "Re-examining Perceived Usefulness and Ease of Use in Online Booking : The Case of Hong Kong Online Users," *International Journal of Contemporary Hospitality Management*, 27 (2), pp. 185-198.

6. Dickinger, A. and Mazanec, J., 2008, "Consumers' Preferred Criteria for Hotel Online Booking," *Information and Communication Technologies in Tourism*, pp. 244-254.

7. Sigala, M., Lockwood, A. and Jones, P., 2001, "Strategic Implementation and IT : Gaining Competitive Advantage from the Hotel Reservations Process," *International Journal of Contemporary Hospitality Management*, 13 (7), pp. 364-371.

8. Sigala, M., 2005, "Integrating Customer Relationship Management in Hotel Operations : Managerial and Operational Implications," *International Journal of Hospitality Management*, 24 (3), pp. 391-413.

9. Xiang, Z., Magnini, V.P. and Fesenmaier, D.R., 2015, "Information Technology and Consumer Behavior in Travel and Tourism : Insights from Travel Planning Using the Internet," *Journal of Retailing and Consumer Services*, 22, pp. 244-249.

10. Yazdanifard, R. and Yee, L.T., 2014, "Impact of Social Networking Sites on Hospitality and Tourism Industries," *Global Journal of Human-Social Science : Economics*, 14 (8).

11. Fuchs, M., Höpken, W. and Lexhagen, M., 2014, "Big Data Analytics for Knowledge Generation in Tourism Destinations – A case from Sweden," *Journal of Destination Marketing and Management*, 3 (4), pp. 198-209.

12. Chekalina, T., Fuchs, M. and Lexhagen, M., 2011, "A Customer-based Brand Equity Model for Tourism Destinations :–Insights from Service Marketing. in Innovation Processes and Destination Development in Tourist Resorts," *The Regional Studies Association Research Network on Tourism and Regional Development Conference*, 30 March–1 April 2011, Östersund, Sweden.

13. Aggarwal, C.C., 2011, "An Introduction to Social Network Data Analytics," pp. 1-15, in C.C. Aggarwal (Ed.) *Social Network Data Analytics*, Springer Science and Business Media

14. Leesa-Nguansuk, S., 2017, Thailand in Social Media World's Top 10 [Online], Available : http://www.bangkokpost.com/ learning/advanced/1255403/thailand-in-social-media-worlds-top-10 [12 July 2017].

15. Charoenphon, V., 2014, The Future of Thailand's Tourism Industry [Online], Available : https://www.scbeic.com/en/detail/product/458 [4 August 2017].

16. Askham, N., Cook, D., Doyle, M., Fereday, H., Gibson, M., Landbeck, U., Lee, R., Maynard, C., Palmer, G. and Schwarzenbach, J., 2013, "The Six Primary Dimensions for Data Quality Assessment," Technical Report, DAMA UK Working Group.

17. Sint, R., Stroka, S., Schaffert, S. and Ferstl, R., 2009, "Combining Unstructured, Fully Structured and

Semi-Structured Information in Semantic Wikis," Sem Wiki, 464.

18. Jain, V., 2007, Big Data and Hadoop, Khanna Publishing.

19. Hurwitz, J., Nugent, A., Halper, F. and Kaufman, M., 2013, Big Data for Dummies, John Wiley and Sons.

20. Finlay, S., 2014, Predictive Analytics, Data Mining and Big Data : Myths, Misconceptions and Methods, Springer.

21. Lucas, J.P., Luz, N., Moreno, M.N., Anacleto, R., Figueiredo, A.A. and Martins, C., 2013, "A Hybrid Recommendation Approach for a Tourism System," *Expert Systems with Applications*, 40 (9), pp. 3532-3550.

22. Yu, H., Jing, Y.D.L. and Mu, Z., 2013, "Research on Personalized Recommender System for Tourism Information Service," *Computer Engineering and Intelligent Systems*, 4 (5), pp. 32-46

23. Loh, S., Lorenzi, F., Saldaña, R. and Licthnow, D., 2003, "A Tourism Recommender System Based on Collaboration and Text Analysis," *Information Technology and Tourism*, 6 (3), pp. 157-165.

24. Magnini, V.P., Honeycutt, E.D. and Hodge, S.K., 2003, "Data Mining for Hotel Firms : Use and Limitations," *The Cornell Hotel and Restaurant Administration Quarterly*, 44 (2), pp. 94-105.

25. Li, G., Law, R., Vu, H.Q., Rong, J. and Zhao, X.R., 2015, "Identifying Emerging Hotel Preferences using Emerging Pattern Mining Technique," *Tourism Management*, 46, pp. 311-321.

26. Juwattanasamran, P., Supattranuwong, S. and Sinthupinyo, S., 2013, "Applying Data Mining to Analyze Travel Pattern in Searching Travel Destination Choices," *International Journal of Engineering Science (IJES)*, 2 (4), pp. 38-43.

27. Pitchayadejanant, K. and Nakpathom, P., 2017, Data Mining Approach for Arranging and Clustering the Agro-Tourism Activities in Orchard [Online], Avaliable : http://www.sciencedirect.com/science/article/pii/ S2452315117300619 [20 August 2017]

28. Rong, J., Vu, H.Q., Law, R. and Li, G., 2012, "A Behavioral Analysis of Web Sharers and Browsers in Hong Kong using Targeted Association Rule Mining," *Tourism Management*, 33 (4), pp. 731-740.

29. Guo, Y., Barnes, S.J. and Jia, Q., 2017, "Mining Meaning from Online Ratings and Reviews : Tourist Satisfaction Analysis using Latent Dirichlet Allocation," *Tourism Management*, 59, pp. 467-483.

30. Rogerson, P., 2001, Statistical Methods for Geography. Sage.

31. Wu, E.H., Law, R. and Jiang, B., 2013, "Predicting Browsers and Purchasers of Hotel Websites : A Weight-of-Evidence Grouping Approach," *Cornell Hospitality Quarterly*, 54 (1), pp. 38-48.

32. Chang, K.C., Chen, M.C., Kuo, N.T., Hsu, C.L. and Cheng, Y.S., 2016, "Applying Data Mining Methods to Tourist Loyalty Intentions in the International Tourist Hotel Sector," *Anatolia*, 27 (2), pp. 271-274.

33. Tama, B.A., 2015, "Data Mining for Predicting Customer Satisfaction in Fast Food Restaurant," *Journal of Theoretical and Applied Information Technology*, 75 (1).

34. Suh, E. and Alhaery, M., 2016, "Customer Retention : Reducing Online Casino Player Churn Through the Application of Predictive Modeling," *UNLV Gaming Research and Review Journal*, 20 (2), p. 6.

35. Pei, J., Han, J., Mortazavi-Asl, B. and Zhu, H., 2000, "Mining Access Patterns Efficiently from Web Logs," *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer.

36. Aldenderfer, M.S. and Blashfield, R.K., 1984, "Cluster Analysis," Sage University Paper Series on Quantitative Applications in the Social Sciences 07-044.

37. Dursun, A. and Caber, M., 2016, "Using Data Mining Techniques for Profiling Profitable Hotel Customers : An Application of RFM Analysis," *Tourism*

Management Perspectives, 18, pp. 153-160.

38. Nasukawa, T. and Nagano, T., 2001, "Text Analysis and Knowledge Mining System," *IBM Systems Journal*, 40 (4), pp. 967-984.

39. Stahl, F., Gaber, M.M. and Adedoyin-Olowe, M., 2014, "A survey of data mining techniques for social media analysis," *Journal of Data Mining and Digital Humanities*.

40. Li, H., Ye, Q. and Law, R., 2013, "Determinants of Customer Satisfaction in the Hotel Industry : An Application of Online Review Analysis," *Asia Pacific Journal of Tourism Research*, 18 (7), pp. 784-802.

41. Schuckert, M., Liu, X. and Law, R., 2015, "Hospitality and Tourism Online Reviews : Recent Trends and Future Directions," *Journal of Travel and Tourism Marketing*, 32, pp. 608-21.

42. Capriello A., Mason P.R., Davis B. and Crotts J.C., 2013, "Farm Tourism Experiences in Travel Reviews : A Cross-Comparison of Three Alternative Methods for Data Analysis," *Journal of Business Research*, 66, pp. 778-85.

43. Stringam, B.B. and Gerdes Jr, J., 2010, "An Analysis of Word-of-Mouse Ratings and Guest Comments of Online Hotel Distribution Sites," *Journal of Hospitality Marketing and Management*, 19 (7), pp. 773-796.

44. Zhang, Z., Ye, Q., Law, R. and Li, Y., 2010, "The Impact of e-word-of-mouth on the Online Popularity of Restaurants : A Comparison of Consumer Reviews and Editor Reviews," *International Journal of Hospitality Management*, 29 (4), pp. 694-700.

45. Park, S.B., Park, S.B., Jang, J., Jang, J., Ok, C.M. and Ok, C.M., 2016, "Analyzing Twitter to Explore Perceptions of Asian Restaurants," *Journal of Hospitality*

and Tourism Technology, 7 (4), pp. 405-422.

46. Sánchez-Franco, M.J., Navarro-García, A., and Rondán-Cataluña, F.J., 2016, "Online Customer Service Reviews in Urban Hotels : A Data Mining Approach," *Psychology and Marketing*, 33 (12), pp. 1174-1186.

47. Liao, S.H., Chen, Y.J. and Deng, M.Y., 2010, "Mining Customer Knowledge for Tourism New Product Development and Customer Relationship Management," *Expert Systems with Applications*, 37 (6), pp. 4212-4223.

48. Yuan, Y., 2015, "The Study on Relationship Verification of e-wom and Hotel Operation Performance by Data Mining Technique," *Journal of Global Economics*, 3 (130), p. 2.

49. Tank, D.M., 2014, "Improved Apriori Algorithm for Mining Association Rules," *International Journal of Information Technology and Computer Science (IJITCS)*, 6 (7), p. 15.

50. Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2016, Data Mining : Practical Machine Learning Tools and Techniques, Morgan Kaufmann.

51. Lakshmi, K., Krishna, M.V. and Kumar, S.P., 2013, "Performance Comparison of Data Mining Techniques for Predicting of Heart Disease Survivability," *International Journal of Scientific and Research Publications*, 3 (6), pp. 1-10.

52. Cankurt, S. and SUBAŞI, A., 2016, "Tourism Demand Modelling and Forecasting using Data Mining Techniques in Multivariate Time Series : A Case Study in Turkey," *Turkish Journal of Electrical Engineering and Computer Sciences*, 24 (5), pp. 3388-3404.

53. Hand, D.J., 2016, "Classifier Technology and the Illusion of Progress," *Statistical Science*, 21 (1), pp. 1-14.