**ISBN** 978-981-11-0008-6
**Proceedings of 2016 6th International Workshop on Computer Science and Engineering (WCSE 2016)**
*Tokyo, Japan, 17-19 June, 2016, pp. 206-213*
doi: 10.18178/wcse.2016.06.033

# To Be or Not Be Competitive Country: Analysis of Travel and Tourism Competitiveness Index by Multiple Data Mining Techniques

Anongnart Srivihok [1] and Arunee Intrapairot [2]

[1] Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
[2] Department of Management, Faculty of Business Administration, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand

**Abstract.** Travel and Tourism Competitiveness (TTC) has been raised as an important issue by World Economic Forum (WEF). The measurements of TTC covering 141 countries around the world provide information for all stakeholders of each country to enhance tourism competitiveness since tourism improvement may lead to increasing national growth and wealth. The purpose of this study is to understand the factors contributing to tourism compeтiveness and their relationships. The Travel & Tourism Competitiveness Index (TTCI) of all countries were collected from WEF reports. Then, the dataset were analysed by three data mining techniques consisting of clustering, classification and association rules mining. The countries are clustered into 8 segments. Characteristics of each cluster and relationships of TTCI are also proposed. The revealed results in this paper can be used by the governments and tourism sectors to develop their strategic plans and management.

**Keywords:** travel and tourism competitiveness, data mining, clustering, classification, association rule mining.

## 1. Introduction

According to the United Nations World Tourism Organization (UNWTO), international tourist arrivals were about 1.14 billion in 2014 with the growth rate of 4.4% to reach a total of 1.18 billion in 2015. In addition, based on the World Travel & Tourism Council (WTTC) data, the travel and tourism sector of the world is about 9.5% of global GDP and 5.4% of world exports, with the total value of US$ 7 trillion. Tourist industry plays an important role as an accelerator of economic growth and employments, growing at 4% in 2014 and touching 266 million of direct and indirect employments. At present, jobs relating to the tourist industry now stands for one in 11 jobs on the world, a number that could even rise to one in 10 jobs by 2022 [1]. Thus, it is important to convince all players of tourism sectors to improve the quality of the tourism industry in every aspect.

According to UNWTO's long term forecast, by 2030 the global arrivals of international tourists are estimated to grow by 3.3% each year from 2010 to 2030 to reach 1.8 billion. From 2010 to 2030, the forecast of international tourist arrivals to emerging destinations (+4.4% a year) are increased two folds comparing to the rate of those in advanced economies (+2.2% a year). As well, the market share of emerging economies rose from 30% in 1980 to 47% in 2013. Furthermore, it is estimated to touch 57% by 2030, with more than 1 billion international tourist arrivals [2].

In 2007, World Economic Forum published first Travel and Tourism Competitiveness (TTC) Report which covered major and emerging economies of 124 countries. Lately, the Reports have been expanded to cover 141 countries in 2015. The TTC Reports also provide the evaluation profiles of each country.

+ Corresponding author. Tel.: +662 5625555 3730; fax: 662 9428488.
*E-mail address*: fsciang@ku.ac.th.

Sources of information are consisted of key economic inducators from the World Economic Forum (WEF). The indicators from the WEF and country indicators measurement of ***Travel and Tourism Competitiveness Index (TTCI)*** include 4 main index 1) Enabling Environment, 2) T&T Policy and Enabling Conditions, 3) Infrastructure and 4) Natural and Cultural Resources. The TTCI is employed to reflect attractiveness for investment in the travel and tourism industry of each country. The measurement via TTCI is better than that of country attractiveness as a tourist destination because TCCI is a global comparative tool accommodating all various aspects of the tourism industry. The selected countries were scored from 1 to 7 according to the performance in each specific TTC sub index.

According to TTC Report 2015, Travel and Tourism Competitiveness Index (TTCI), the four main indexes are divided into 14 pillars. First, Enabling Environment includes 5 pillars: 1) Business Environment, 2) Safety and Security, 3) Health and Hygiene, 4) Human Resources and Labour Market and 5) ICT Readiness. Second, T&T Policy and Enabling Conditions are supported by 4 pillars including 6) Prioritization of Travel and Tourism, 7) International Openness, 8) Price Competitiveness, and 9) Environmental Sustainability. Third, Infrastructure includes three pillars of 10) Air Transport Infrastructure, 11) Ground and Port Infrastructure, and 12) Tourist Service Infrastructure. Fourth, Natural and Cultural Resources includes two pillars: 13) Natural Resources and 14) Cultural Resources and Business Travel. The scale of TTCI is ranging from 1 to 7, where 1 is the lowest and 7 is the highest score [1].

## 2. Prior Studies

### 2.1. Data Mining Techniques

Three data mining techniques are applied in this study consisting of clustering, classification and association rule mining.

### 2.2. Data Mining in Tourism

Data mining techniques were applied to study tourism data in East Asia by Guoxia and Jianqing [3]. Tourist dataset were obtained from Guilin, China and analysed using a classification algorithm (decision tree-C4.5). The study reported that the shopping environment was associated with the tourists' impressions. However, tourism products focusing on the attractiveness of touristic places would not relate to tourist satisfaction.

In 2009, the Chinese researchers Kou, Wang, Hwang and Ye [4] proposed a relationship between tourism demand and the exchange rate by applying a data mining technique to tourism data from 8 Asian countries consisting of Japan, China, South Korea, Taiwan, Hong Kong, Singapore, Malaysia, and Thailand. Tourism demand was measured by the numbers of inbound tourists. The results revealed that the exchange rate influenced tourism demand in Asian countries. The lower the exchange rates in the target countries for tourists, the higher the number of tourists and the more support for the tourism industry.

Five different classification techniques—Bayes Network, Radial Basis Function, Pruned Tree, Single Rule Learner and Nearest Neighbours algorithms—were compared for their performance on a breast cancer data set by Othman and Yau [5]. According to the comparison results, the best classifier for the given data set was a Bayesian Network with a high accuracy rate of 89.71% and the lowest average error at 0.2140. The results indicated that among the five classifiers, a Bayesian Network had a high possibility of enhancing the classification algorithm for using in general cases.

Yotsawas and Srivihok [6] presented inbound tourist segmentation with coupling algorithms using K-Means and Decision Tree. The study was divided into two phases: clustering phase, and classification phase. First, the segmentation was carried out by Self Organizing Map (SOM) and K-Means. SOM was used to find the number of cluster. Next, K-Means algorithm was applied for segmentation. Second, three classifiers were compared on the performances of prediction accuracy. Three classifiers consist of Decision Tree, Naïve Bayes and Multilayer Perceptron (MLP). The predictive ability of J48 Decision Tree was the best of three classifiers with the accuracy as 99.57%. Recently, an inbound tourist data set was mined by Srivihok and Intrapairot [7]. Featured selection and classification were applied to obtain the production rules. In total, there were 14 tourist attributes. After feature selection using a Consistency Feature Selection Subset Evaluator Algorithm and the Best First Searching Method, seven attributes were extracted (i.e. gender, age, income, occupation, purpose, domestic air transportation, and car transportation). Then, these seven features were used for classification using REPTree. The study supported the claim that feature selection improved the accuracy of classification.

The objective of this study is to segment the country competitiveness in clusters according to their TTCI, and further find the relationship of TTCI with their cluster and among themselves.

# 3. Methodology

Data mining software, Weka version 3.7 was used for data analysis which included four steps: 1) data pre-processing 2) data clustering, 3) data classification, and 4) association rule mining.

## 3.1. Data Collection

The data set of Travel and Tourism Competitiveness Index (TTCI) from 141 countries obtained from UNWTO reports from year 2007, 2008, 2009, 2011, 2013 and 2015 were combined to one data set [1, 2, 8]. Then, incomplete records and missing data were deleted. Finally, 805 instances were selected. At present the TTCI was adapted from 14 indexes to 13 indexes since the last two indexes: no. 13. Natural resources and no. 14: Cultural Resources are combined to one index named no. 13: Natural/Cultural resources. Thus, 13 indexes are applied in this study including 1) Business Environment, 2) Safety and Security, 3) Health and Hygiene, 4) Human Resources, 5) ICT Infrastructure, 6) Prioritization of Travel & Tourism, 7) International Openness 8) Price Competitiveness in the T&T Industry 9) Environmental Sustainability, 10) Air Transport Infrastructure, 11) Ground Transport Infrastructure, 12) Tourism Service Infrastructure, and 13) Natural and Cultural Resources.

Table 1 shows the analysis of TTCI scores, ranged from 1 to 7 (i.e. 1= lowest score, and 7= highest score). Each country was assigned TTCI scores for 13 indexes as indicated on the above paragraph.

Table 1: TTCI of Each Country

| **Country name** | **Index** | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **.** | **13** |
| Albania | 4.14 | 2.5 | 4.09 | 4.81 | . | 3.68 |
| Algeria | 3.37 | 3.66 | 4.18 | 4.91 | . | 4.28 |
| Angola | 2.93 | 2.92 | 3.85 | 2.61 | . | 3.02 |
| Argentina | 4.30 | 3.41 | 3.90 | 4.04 | . | 4.93 |
| Armenia | 3.81 | 3.28 | 4.79 | 5.46 | . | 3.98 |
| Australia | 4.81 | 5.58 | 5.50 | 5.91 | . | 5.64 |
| Austria | 5.33 | 6.09 | 6.20 | 6.18 | . | 6.52 |
| Bahrain | 4.71 | 3.74 | 4.55 | 4.76 | . | 4.27 |

## 3.2. Research Framework

The study framework is depicted in Fig. 1. After the TTCI dataset is cleaned and preprocessed. The remaining usable data is about 805 records. Then, dataset was preprocessed by digitization to transform numeric data to nominal data. For all data analysis in this study Weka 3.7.11 open source software [9] was used.

***Clustering*** is applied to segment the dataset according to common structures or patterns of attributes. Expectation Maximisation (EM) clustering algorithm [10] is applied for segmenting TTCI dataset (Table 1). Then, it results in new dataset which cluster numbers are assigned in each record named TTCIClus (Table 2).

***Classification*** is the supervised learning algorithm as the analysis of known structures to new data into different classes. This analysis is performed by using classification model.

The experiments were designed as follows. TTCIClus dataset was analysed with three types of classification algorithms: Bayes, rule base and decision tree. Bayes includes Bayes Net and Naïve Bayes. Rule base consists of JRIP and PART, while decision tree includes J48 or C45, and CART [10]. In this analysis, cluster number was used as class label (Table 2). Thus, data set was tested with the above six classifiers. Ten folds cross validation were applied for training and testing classification model.

***Association Rule Mining*** is the technique to find the relationships of attributes in the dataset. In this study Apriori algorithm is applied.

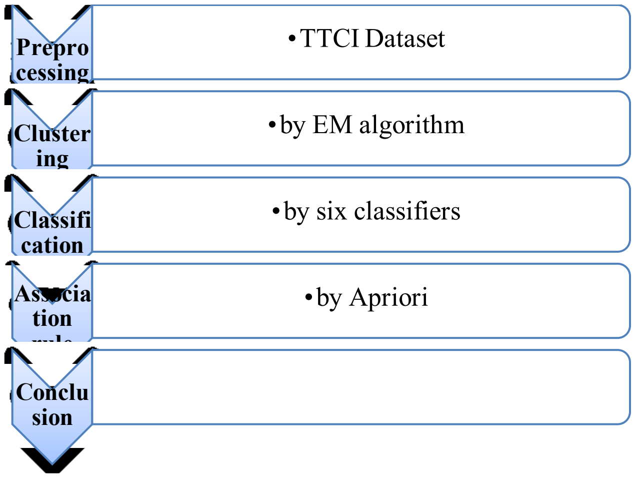


Fig. 1: The study framework.

## 4. Experiments

### 4.1. Data Clustering

EM clustering algorithm has been applied to cluster TTCI dataset containing 805 records of 141 countries reporting by WEF from year 2007, 2008, 2009, 2011, 2013, and 2105. TTCI dataset was segmented into 8 clusters. Seven clusters are almost the same sizes which are ranging from 11%-17% while Cluster 6 is the smallest cluster about 2% (Fig. 2). Each record in TTCI dataset was assigned its cluster number as a new item as depicted in Table 2. This new dataset name is TTCIClus.

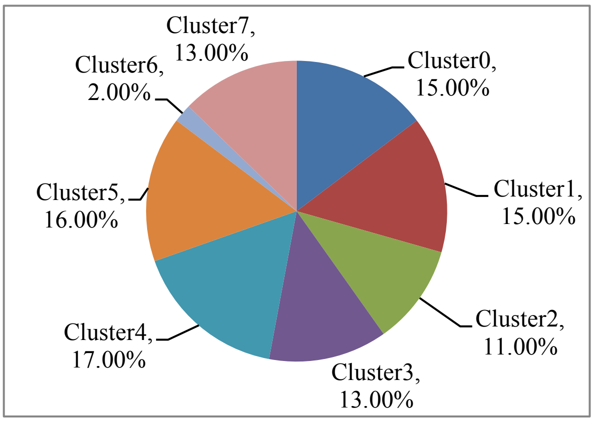


Fig. 2: Clustering data with EM algorithms into 8 clusters.

Table 2: TTCIClus Dataset Generated by EM Clustering

| **Country** | **Ind1** | **Ind2** | **Ind3** | .. | **Ind12** | **Ind13** | **Cluster** |
|---|---|---|---|---|---|---|---|
| Algeria | 3.78 | 4.9 | 4.97 | .. | 2.04 | 2.05 | Cluster0 |
| Argentina | 2.67 | 5.03 | 6.2 | .. | 4.08 | 4.37 | Cluster1 |
| Armenia | 3.42 | 5.8 | 5.92 | .. | 1.88 | 1.38 | Cluster0 |
| Australia | 5.04 | 6.24 | 6.14 | .. | 5.31 | 5.13 | Cluster7 |
| Austria | 4.94 | 6.47 | 6.97 | .. | 4.07 | 2.92 | Cluster7 |
| Azerbaijan | 3.48 | 4.45 | 5.83 | .. | 2.03 | 1.43 | Cluster0 |
| Bahrain | 5.53 | 5.33 | 5.17 | .. | 1.92 | 1.33 | Cluster2 |
| Bangladesh | 4.06 | 4.43 | 4.29 | .. | 2.3 | 1.56 | Cluster3 |
| Barbados | 4.62 | 5.75 | 6.02 | .. | 2.17 | 1.13 | Cluster0 |
| Belgium | 4.71 | 6.18 | 6.49 | .. | 2.65 | 3.67 | Cluster7 |

Table 3: The Characteristics of Cluster 0 to Cluster 7 Segmented by Using EM Algorithm

| **TTCI index** | **C0** | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** | **C7** |
|---|---|---|---|---|---|---|---|---|
| 1. Business Environment | l | m | l | l | m | l | h | h |
| 2. Safety & Security | l | m | m | h | h | l | h | h |
| 3. Health & Hygiene | l | m | l | h | h | l | h | h |
| 4. Human Resources | l | m | l | m | h | m | h | h |
| 5. ICT Infrastructure | l | m | l | l | h | l | h | h |
| 6. Prioritization of Travel and Tourism | l | m | l | m | m | m | h | h |
| 7. International Openness | l | l | l | l | m | l | m | h |
| 8. Price Competitiveness | h | h | h | m | m | m | h | l |
| 9.Environment Sustainability | l | l | l | l | m | l | l | h |
| 10. Air Transportation Infrastructure | l | m | l | l | m | m | h | h |
| 11. Ground and Port Transportation Infrastructure | l | l | l | l | h | l | h | h |
| 12. Tourism Service Infrastructure | l | m | l | m | m | l | m | h |
| 13. Natural and Culture Resources | l | l | l | l | m | l | m | h |
| Cluster Name | LC | PC | LC | AP | AP | PC | PC | HE |

Note: C0-C7= Cluster0-Cluster7, l= lower than average score, m= medium or average scores, h= higher than average scores, AV= average scores, LC= low cost, PC= price competitiveness, HE=high end.

The characteristic of each cluster is depicted in Table 3. Then, clusters are named according to the score of main characteristics.

**Cluster 0** is named as ***Low Cost*** since almost all attributes scores are low except attribute *Price Competitiveness*. Countries in this cluster are developing countries.

**Cluster 1** is named as ***Price Competitiveness*** with many average scores of TCCI.

**Cluster 2** is also ***Low Cost*** with medium safety and security.

**Cluster 3** is ***Average Price Competitiveness*** with high *Safety and Hygiene*.

**Cluster 4** is also ***Average Price Competitiveness*** with high *Safety and Security, Human Resources and ICT Infrastructure*.

**Cluster 5** is an ***Average Price Competitiveness***.

**Cluster 6** is ***High Price Competitiveness*** with many high scores in many attributes.

**Cluster 7** is ***High End*** which all attribute scores are high except *Price Competitiveness*.

## 4.2. Data Classification

TTCIClus dataset (as depicted in Table 2) were used for generating classification models by six classifiers: CART, J48, JRip, PART, Naïve Bayes and Bayes Net.

Table 4: Accuracy of Six Classifiers in Classification of TTCIClus Dataset

| **Tree** | | **Rule** | | **Bayes** | |
|---|---|---|---|---|---|
| ***CART*** | ***J48*** | ***JRip*** | ***PART*** | ***NB*** | ***Bayes Net*** |
| 81.84% | 74.13% | 70.77% | 73.26% | **97.39%** | **97.14%** |

Results from Table 4 show that Naïve Bayes (NB) and Bayes Net are the two classifiers having the highest accuracy rates of 97.39% and 97.14%, respectively. While, two decision tree classifiers, CART and J48, provided 81.84% and 74.13% accuracy rate. Last, JRip and PART are the least performance with the lowest accuracy rate of 70.77% and 73.26%, respectively. Accordingly, with six different classifiers it is likely that 13 TTCI are good factors to build model for predicting the class (cluster number).

## 4.3. Association Rule Mining

After applying Apriori algorithm to TTCIClus dataset with Minimum support= 0.06 and Minimum metric <lift>= 0.1, 100 rules are generated. Some redundant rules are eliminated. Then the significant rules are as follows:

Table 5: Association Rules of TTCIClus Data Set Generated by Apriori Algorithm

| |
|---|
| 1) ICT infrastructure <=1.7 ==> Cluster Num=cluster0<br>2) Natural and cultural resources <=1.5==> Cluster Num=cluster0<br>3) Air transport infrastructure =1.9-2.5==> Cluster Num=cluster0 |
| 4) Air transport infrastructure =3.5-4.0 ==> Cluster Num=cluster1<br>5) Tourism infrastructure=3.4-4.0==> Cluster Num=cluster1<br>6) Health and hygiene=4.6-5.2 ==> Cluster Num=cluster1 |
| 7) Ground transport infrastructure = 2.1-2.6 == > Cluster Num=cluster2 |
| 8) Ground transport infrastructure= 4.4-5.2==> Cluster Num=cluster4<br>9) Health and hygiene = 5.8-6.4 == > Cluster Num=cluster4<br>10) Safety and security = 5.4-5.9 == >Cluster Num=cluster4 |
| 11) Tourism infrastructure = 2.2-2.8 ==> Cluster Num=cluster5 |
| 12) Price competitiveness = 3.4-3.8 ==> Cluster Num=cluster7<br>13) Policy rules and regulations= 5.0-5.05 ==> Cluster Num=cluster7<br>14) Safety and security = 5.8-6.3 == >Cluster Num=cluster7 |
| 15) Policy rules and regulations = 5.0-5.4 == > Human resources = 5.3-5.8<br>16) Tourism infrastructure <= 1.6 ==>Natural and cultural resources =< 1.6<br>17) ICT infrastructure<= 1.7 == > Natural and cultural resources <=1. 6<br>18) Tourism infrastructure =1.6-2.2 ==> ICT infrastructure = 1.7-2.2 |

Association rules from Table 5 depicted the relationships of TTCI and Cluster number. Countries which are in Cluster 0 have low scores of ICT and Transport Infrastructure and also Natural Cultural Resources as well. Cluster 1, three indexes: Tourism Infrastructure, Air Transportation Infrastructure, and Health and Hygiene are about average. Cluster 2, Ground Transportation score is below average. Cluster 4, all three indexes scores: Ground Transportation Infrastructure, Health and Hygiene, and Safety and Security are above average. Cluster 5, one index, Tourism Infrastructure score is lower than average. Last, Cluster 7, Price Competitiveness is lower than average. On the contrary, Policy Rules and Regulation and Safety and Security are higher than average.

According to rules 15 to 18, there are some associations between TTCI. Countries which have High Policy Rules and Regulations always have high Human Resources (rule 15). Some pairs of index indicate low scores, such as Tourism Infrastructure and Natural and Cultural Resources scores (rule 16), ICT Infrastructure and Natural and Cultural Resources (rule 17), Tourism Infrastructure and ICT Infrastructure (rule 18), and Air Transport Infrastructure and Tourism Infrastructure (rule 18).

## 5. Conclusions

Travel & Tourism Competitiveness Index (TTCI) has been accepted as significant factors which contribute to improving quality of tourism sectors leading to increasing the wealthy and sustainability of the countries. This study investigated the scores of TTCI in 141 countries around the world from 2007-2010, 2011, 2013, and 2015. Three data mining techniques consisted of clustering, classification and association rule mining were applied to analyse the TTCI dataset.

Clustering TTCI dataset by means of EM algorithm results in 7 clusters of countries. Each cluster has distinct characters. For example, Cluster 7 named as 'Hi End', the members of this cluster are developing countries with highly competitive in information technology and air and ground transport infrastructure such as Switzerland and Austria. However, the price competitiveness score of this cluster is the lowest. It implies that travelling to the countries in this cluster is expensive. Cluster 0 named as 'Low Cost', members of this cluster are developing countries. All TTCI scores are very low except the price competitiveness. The results indicate that low qualified products and services may lead to being low cost destinations for tourists' perception. There are some clusters which are between the LOW COST and Hi END (i.e. Cluster 1, 2, 3, 4, 5 and 6).

Next step is data classification, six different classifiers consists of Bayes Net and Naïve Bayes, JRIP and PART, J48 or C45, and CART were used to generate the classification models from TTCIClus dataset. The performances of class prediction were tested. Results show that Bayes Net and Naïve Bayes classifiers are the best with high prediction accuracy rates of 97.14% and 97.39% respectively. It seems that all 13 TTCI and the former two classifiers are good components in building classification models.

It is questionable that any of TTCI is related to each other, so the last step is to find relationship by association rule mining. Thus, Apriori algorithm was applied to generate the association rules from TTCIClus dataset. Generated rules show the relationships of some TTCI and its cluster number and also the relationships among themselves. Likely, some rules obtained from this step are close to the cluster characteristics summarized from clustering. The interesting association rules such as "countries with low score in ICT infrastructure also have low score in tourism infrastructure". As well, results from clustering dataset show that countries in Cluster 0 have low scores in both tourism and ICT infrastructure.

If countries in Cluster 0 demand to increase their competitiveness in travel and tourism they should emphasize on developing strategic plans and improving their infrastructures including air and ground transportation, ICT and tourism services. On the contrary, countries in Cluster 7, HI End, which are highly competitive, should reconsider on pricing. If the countries aim at decreasing the prices to allow for more tourists they can employ modern technology to decrease costs of travelling and services. However, if cost focus is not the main considerations, they are able to customize tourism packages to fulfil their tourists' expectations.

In addition, the study revealed that TTCI can be used to improve the drawbacks of tourism attributes. For example, Thailand one of the "*Price Competitiveness*" countries has to resolve the problems relating to ICT Infrastructure and Safety & Security. Yet, all TTCI can also be employed to benchmark with other neighbouring countries not only for competition but also for supporting up-selling and cross selling marketing.

With the three data mining techniques, it is likely that TTCI scores have been analysed thoroughly and accurately and provide the consistence results. Information provided by this study can be applied in the policy and regulation rules formation to improve the country competitiveness.

## 6. Acknowledgement

This study was supported by Department of Computer Science, Kasetsart University, Bangkok, Thailand.

## 7. References


[1] *The Travel and Tourism Competitiveness Report 2015*. www.weforum.org/reports/travel-tourism-competitiveness-report-2015. Date 1 February 2016.

[1] *Tourism Towards 2030 / Global Overview - Advance edition presented at UNWTO 19th General Assembly* - 10 October 2011. UNWTO Tourism Highlights 2014 Edition. http://mkt.unwto.org/publication/unwto-tourism-highlights-2014-edition. Date 1 February 2016.

[2] Z. Guoxia, and T. Jianqing. The application of data mining in tourism information. *IEEE International Conference on Environmental Science and Information Application Technology*, (2009) pp. 689-692.

[3] H. Kou, H. Wang, W.Y. Hwang, and C.Ye. Tourism Demand and Exchange Rates in Asian Countries: Evidence from a Panel Data Approach. *IEEE World Congress on Computer Science and Information Engineering*, 2009, 4, 508-512.

[4] M. F. Othman, and T. M. S. Yau. Comparison of Different Classification Techniques Using WEKA for Breast Cancer, *3rd Kuala Lumpur International Conference on Biomedical Engineering. IFMBE Proceedings*, 2007, 15, 520-523.

[5] V. Yotsawas and A. Srivihok. Inbound Tourists Segmentation with Combined algorithms using K-Means and Decision Tree. *Proceedings of the 2013 10$^{th}$ International Joint Conference on Computer Science and Software Engineering. JCSSE 2013*. 189-194.

[6] A. Srivihok and A. Intrapairot. Enhanced Classification Performances of Travel and Tourism Competitiveness Model with Feature Selection. *Advanced Science Letter*. 2015, 21 (10): 3170-3174.

[7] *The Travel & Tourism Competitiveness Report 2011 Beyond the Downturn*, 2011. www.weforum.org/reports/ travel-tourism-competitiveness-report-2011. Date 1 February 2016.

[8] *Weka Software Version 3.7.11*. http://www.cs.waikato.ac.nz/ml/weka/. Date 1 February 2016.

[9] I. H. Witten, and E. Frank. *Data Mining, Practical Machine Learning Tools and Techniques*. Second Edition. Morgan Kaufmann Publishers, 2005.