



Data mining approach for arranging and clustering the agro-tourism activities in orchard



Krittipat Pitchayadejanant ^{a,*}, Parinya Nakpathom ^b

^a Business Data Analysis, International College, Burapha University, Chonburi 20131, Thailand

^b Department of Hospitality and Tourism Management, International College, Burapha University, Chonburi 20131, Thailand

ARTICLE INFO

Article history:

Received 28 February 2017

Received in revised form 24 May 2017

Accepted 5 July 2017

Available online 1 August 2017

Keywords:

agro-tourism activities,
association rule,
clustering,
data mining,
market basket analysis

ABSTRACT

The paper attains to finding the association among interested activities in orchard that stimulate the tourists to travel. The knowledge obtained in this study is applying data mining techniques to create the association rule in order to find out the pattern of activities for orchard tourism. The evoked set of activities is the most frequent set which travelers would like to do when they visit orchard. The tool of analyzing the association rules is Rapid Miner 7.3. The result shows that the highest recommended activity is reaping and tasting the fruit. Another activities, which also are recommended to be arranged more into orchard, consist of walking, shopping, and feeding animal in the orchard.

© 2017 Kasetsart University. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Agriculture is essence and be the basis of Thailand's economy. Although, Thailand is the top rank of export the agriculture products, the poverty of farmers is still high. The agriculture in the country is uncertain because of climate change, high competition in ASEAN countries, fallacy from the government's support, and etc. However, the tourism industry is still blossom and should be adapted by integrating agriculture into tourism industry to reduce the poverty of farmers. In addition, Thailand has many fascinated destinations and been developed to enhance the reputation for being the high quality of travel destination.

Thailand has the outstanding identity in culture that attracts the local and foreign tourists. Agro-tourism in Thailand has been promoted by Tourism Authority of Thailand; TAT because Thailand is recognized as the one of agricultural country with abundance of natural resources. The way of agriculturists' life is one of traditional culture that can be adopted and supported in tourism. Philosophy of Sufficient Economy by his Majesty King Bhumibol Adulyadej has been adopted for their living and being survive in capitalism. Hence, agro-tourism is introduced and be interested because it offers the new experience to travelers.

Normally, the travelers usually ask what the interesting activities to do when they arrive at a particular destination. Activities are the link between tourist motivations and destination choices (Moscardo, Morrison, Pearce, Lang, & O'Leary, 1996). Attractive set of activities is the product characteristic that the customer will purchase as being as the product component in marketing. The travelers will be interested and motivated when choices of exotic activities

* Corresponding author.

E-mail address: krittipat@scitech.au.edu (K. Pitchayadejanant).

Peer review under responsibility of Kasetsart University.

information from a certain destination have been heard by travelers.

In tourism industries, Data Mining is applied to analyze the customers widely: forecasting expenditures of tourists, analyzing profiles of tourists, and forecasting number of tourist arrivals (Bose, 2009). The aggressive of competition leads businesses in hospitality and tourism adapt themselves to understand customers for attaining customer loyalty. Notwithstanding, data mining is the process of extracting useful information to analyze the customer behaviour for increasing the quality of services and profitability. Its process is combined by using statistical methods, mathematics, machine learning and scientific methodologies (Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998). Several core techniques that are applied to describe the customer behaviour consist of classification, segmentation, association, prediction and etc. (Girish & Promila, 2016).

Association analysis is the technique in data mining to discover the relationships among the attributes in data set. It is essential technique and interested in retail business (Brijs, Swinnen, Vanhoof, & Wets, 1999), E-Commerce (Chajri & Fakir, 2014), even in education (Merceron & Yacef, 2008). Retailers can find out the association of purchasing one product to another items of customers in order to launch the campaign or promotion to stimulate customers to purchase the product more than one product. With the classic example in data mining, there is high association of purchasing diapers and beer of husband in a large supermarket chain in USA.

Consequently, the study aims to find out the set of interested and attractive activities that the tourists would like to visit an orchard by using data mining techniques. The pattern of these activities will be suggested when an orchard is developed to be tourist destination. In addition, the set of activities will be classified in order to segmenting the visitors. The marketers would be better understanding for matching the activities into different clusters. They will be convenient to provide accurate information to motivate the visitors to visit orchard. The benefit of this study is able to suggest for doing marketing plan in order to promote the orchard as tourist destination for attracting the interest of travelers who will travel in an orchard.

Literature Review

Agro-tourism is promoted based on the relation between agriculture and tourism. This tourism is opened to travelers who seek for nature, entertainment, cultural education, and other activities, but the farm's owner can also generate more income through visitors rather than only the harvest (Petroman & Petroman, 2010). Agro-tourism provides various activities such learning way of life and the culture of communities. There are many activities to support the agricultural; fruit and vegetable picking, testing bee honey, deciphering the craft of wine-making, purchasing souvenirs from souvenir shops, accommodation, food, event-festivals, and etc (Petroman & Petroman, 2010).

In tourism, the characteristics of product are tour program or attractive activities. The destination choice will provide the connection between travelers and destinations.

The activities of destination can segment the travelers who satisfy to travel in a certain destination (Moscardo et al., 1996). Moscardo et al. (1996) classify the travelers by travel benefit into three clusters: social status, self-development, and relaxation. Furthermore, they investigate whether what the activities are preferred in each cluster.

Thailand Institute of Scientific and Technological Research; TISTR classified the activities of agro-tourism into six groups: 1) products display or demonstration such as display the steps of growing rice, 2) agricultural activities by tourists' participations such as fruit gathering, 3) home stay, 4) agricultural knowledge training which provides agricultural knowledge and the wisdom of villagers such as fruit preservation, 5) agricultural products distribution such as selling flowers, fresh fruits, and 6) agricultural business offering.

The activities in this research are listed according to TISTR framework and the activities arranged currently by orchard farm in the East of Thailand which are categorized as shown in Table 1.

Data Mining Techniques in Hospitality and Tourism Application

Data mining techniques are widely applied into hospitality and tourism industry especially to manage the relationship with customers. There are many techniques and algorithms applied such as clustering (Dolnicar, 2002; Lau, Lee, & Ho, 2005; Liao, Chen, & Deng, 2010; Segall & Zhang, 2008), regression (Yuan, 2015), association (Karathiya, Sakshi, Sakshi, & Kathiriya, 2012; Liao et al., 2010), decision trees (Karathiya et al., 2012; Tama, 2015), forecasting (Au & Law, 2002) and etc. The technique to be applied based on data characteristics and business requirements (Giraud-Carrier & Povel, 2003).

Customer relationship management; CRM is the imperative strategy to attain the customers' loyalty. Data mining techniques are implemented to manage the relationship with customers. These techniques are the tools to understand and predict the information of customers such as customers' profiles, customers' behaviour, or customers' needs and wants. Liao et al. (2010) apply two data mining

Table 1
List of orchard activities matching between TISTR and this study

TISTR	List of orchard activities
1 Products display or demonstration	1 Planting 2 Learning agricultural product transformation
2 Agricultural activities by tourists' participations	3 Fruit reaping and tasting 4 Walking in the orchard 5 Bicycling in the orchard 6 Attending festival 7 Feeding animal
3 Homestay	8 Homestay
4 Agricultural knowledge training	9 Learning to do agricultural
5 Agricultural products distribution	10 Learning farmer's lifestyle 11 Purchasing souvenir
6 Agricultural business offering	12 Purchasing sapling or seed 13 Business contact

techniques: clustering and association rules for discovering what characteristics of new product should be developed and which segment was matched with a particular new product.

Even in fast-food restaurant, data mining technique is implemented to predict the customer satisfaction (Tama, 2015). His study is to identify determinant factors that influence customer satisfaction. Customer satisfaction is the crucial determinant that leads to customer loyalty. He applied two algorithms: C4.5 and REANN to generate the rule in order to determine the crucial factors as being as significant predictor of customer satisfaction. Keyword analysis is a technique in data mining. The objective of the technique is to gain insight interconnection between various keywords. Bach, Schatten, and Marusic (2013) apply keyword analysis in data mining to visualize and investigate the application of data mining in tourism publication (Bach et al., 2013). They found that the popularity of data mining techniques in tourism were segmentation, neural networks and forecasting which were benefit for doing marketing.

In addition, web mining is applied to investigate the customers' opinions and comments toward hotel's quality (Segall & Zhang, 2008). The data mining techniques applied in text mining consist of cluster, link analysis, keywords analysis, taxonomy and dimension matrices. The benefit of text mining is that the hotel manager could analyze the information to improve the quality of service in hotel. He/she understands the customer in each segment whether what they need. The information of each segment is analyzed particularly for attaining customers' loyalty.

Yuan (2015) applies data mining technique: regression analysis to study relationship of electronic word of mouth; e-WOM and hotel operation performance (Yuan, 2015). The indicators of hotel operation were categorized into 8 factors: 1) booking and check-in services, 2) price, 3) location, 4) services, 5) business services, 6) sanitary and cleanliness, 7) environmental comfortable, and 8) sleeping quality. The result shows positive relationship with e-WOM and hotel operation performance. It implies that the higher internet exposure will impact the performance of hotel operation positively. Hence, hotel manager should investigate and studied the comments of customers through online in order to address the problems instantaneously.

Based on literature review, this study focuses two data mining techniques: association rule and cluster analysis in order to obtain the information for response the objective of study.

Association Rule

Association rule algorithm is to establishing the correlation between one items and another items. The patterns of item set are evoked. Association rule algorithms that are currently used and be well known are Apriori Algorithm and FP-Growth Algorithm (Tan, 2006). There are seven differences between these two algorithms and their performances were evaluated (Mythili & Shanavas, 2013). Nevertheless, the result shows the performance of FP-Growth be better than Apriori Algorithm: consuming less memory and less time spending.

FP-Growth Algorithm

FP-Growth stands for Frequency Pattern. The algorithm of FP-Growth starts with the frequent patterns 1-itemset and grows in each itemset by its conditional pattern-base: minimum support. However, to obtain the best rules in association rule, there are three thresholds: the minimal support, minimal confidence, and lift. The finding of rules should have the values higher than thresholds which are specified by researcher (García, Romero, Ventura, & Calders, 2007). They need to be given the appropriate parameters for obtaining the suitable rules.

Support is measuring how the items occur frequently which is measured by

$$\text{support}(X) = \frac{\text{count}(X)}{N}$$

The high value of support shows the repeated item frequently which it means the item is selected more often. The aspect to select the itemset should have high support value. The itemset which has low support is concerned as uninteresting rules (Tank, 2014).

Confidence is the measurement of predictive power or its accuracy which is measured by

$$\text{confidence}(X \rightarrow Y) = \frac{\text{support}(X, Y)}{\text{support}(X)}$$

The value of confidence ranges between 0 and 1. The association with the highest confidence should be considered because it shows more reliable (Azevedo & Jorge, 2007). However, confidence solely may not be enough to assess or describe the result (North, 2012). In addition, high confidence rules can be misleading because it ignores the support itemset in the rule (Tan, 2006). Lift value is applied to address the problem.

Lift is the measurement of how much more likely one item is to be selected relative to its typical select, given that another item has been selected. Its value is calculated by

$$\text{lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{support}(Y)}$$

However, $\text{lift}(X \rightarrow Y)$ is the same as $\text{lift}(Y \rightarrow X)$.

The lift value equals to 1 implied that X and Y are independent. If the lift value is below 1, it implies that X and Y are negatively correlated. For the lift value exceeds 1, it presents that X and Y are positively correlated. Hence, the threshold lift value should be greater than 1; it reflects that the two items are found together more often than by chance which it can be implied that a true connection between the items (Lantz, 2013).

This study also would like to do cluster analysis in order to segment the visitors based on activities. The advantage of doing segmentation based on activities is understanding what activities are interested in each segment.

Cluster Analysis

Cluster is a technique that attempts to group the observations which have similar characteristic to find out valuable comprehensive information in each group

Table 2

Description of attributes used for applying in FP-Growth Algorithm

Activities	Possible values
Fruit reaping and tasting (reap and taste)	1 = Yes, 0 = No
Walking in the orchard (walking)	1 = Yes, 0 = No
Bicycling in the orchard (bicycling)	1 = Yes, 0 = No
Learning to do agricultural (learning agriculture)	1 = Yes, 0 = No
Attending festival (festival)	1 = Yes, 0 = No
Feeding animal (feeding)	1 = Yes, 0 = No
Purchasing souvenir (shopping)	1 = Yes, 0 = No
Purchasing sapling or seed (buy seed)	1 = Yes, 0 = No
Planting (planting)	1 = Yes, 0 = No
Learning farmer's lifestyle (learning lifestyle)	1 = Yes, 0 = No
Learning agricultural product transformation (transform)	1 = Yes, 0 = No
Business contact (business)	1 = Yes, 0 = No
Homestay (homestay)	1 = Yes, 0 = No

(Dolnicar, 2002). Cluster analysis is used to do market segmentation. In marketing, doing segmentation aims to launch the advertising, public relation, or plan integrated marketing communication to the direct targeted segments.

In order to gaining the sufficiently distinguished cluster; using only demographic might not be effective (Pesonen, 2013) to classify heterogeneity. Integrating more information of consumer behaviour would be benefit. Pesonen (2013) contributed to increase the segment heterogeneity by using activity and travel motivation. The result shows that the activity based produces more heterogeneity of segmentation. The recommendation system to suggest the travel products which is traveling activities will be recommended to travelers.

There are two techniques of clustering which are widely used: hierarchical clustering and k-means clustering. This study focuses on hierarchical clustering with cluster dialog on binary variable-based distance measurement because the data are binary data.

Research Methodology

Data Understanding and Data Collection

The data are collected through questionnaire: what are activities you would like to do when you travel in orchard.

Multiple response question is applied into this question. The attribute of research is the activities that he/she would like to do in orchard with the value between 1 and 0 whereby the value 1 represents the tourist is interested whereas 0 represents the tourist is not interested to do the activity as shown into Table 2. The questionnaires are distributed to the person who intends to travel at orchard or have ever been traveling in orchard and lives in Bangkok and metropolis or East of Thailand. The data are collected by either survey or email. The sample of 409 observations is analyzed.

Data Preparation

After the questionnaires are collected, the process of data preparation is resumed. The data are recorded into spreadsheet file such Excel. The value in each attribute is recorded between 0 and 1. Those activities are interested to join will be recorded as 1. Otherwise, it is recorded as 0. Rapid Miner 7.3 is used to analyze in FP-Growth Algorithm. The Excel data file is compatible with Rapid Miner 7.3.

Modelling in Association Rules

This research emphasizes on three measurements' criteria: support, confidence and lift.

Building a Set of Rules with the Association Rule Mining Principle

The procedures to create the rules in this study have two phrases:

- 1) Identification of all itemsets with the threshold of minimum support. This study, the threshold of minimum support is defined as 0.200.
- 2) Creating the rules from itemsets which meet the threshold of confidence and lift. The threshold of minimum confidence in this study is identified as 0.700 whereas the threshold of minimum lift is 1.00.

The rules that do not meet the above mentioned minimum thresholds are eliminated (Figure 1).

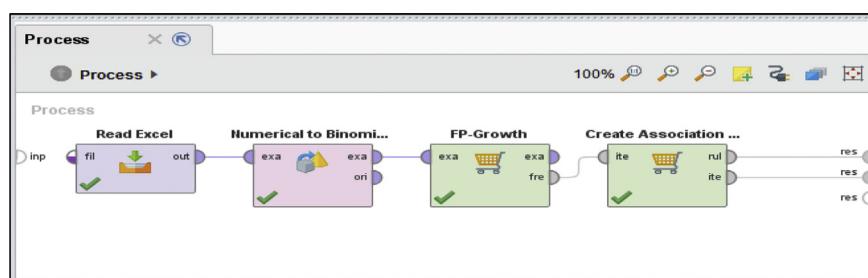


Figure 1 Schema of the rule discovery process in Rapid Miner 7.3

Table 3
Descriptive information of observations

Categories		Frequency	Percent
Gender	Male	159	38.87
	Female	250	61.13
	Total	409	100.0
Marital Status	Single	297	72.61
	Married	104	25.43
	Divorce/Widow	8	1.96
Traveling to orchard with	Total	409	100.0
	Family	162	39.60
	Friends	197	48.17
	Club/Organization/Foundation	46	11.25
	Individual	4	0.98
	Total	409	100.0

Data Analysis and Results

The descriptive information about the characteristics of observation is shown into Table 3. The majority of respondents' gender is female. The marital status is single. The majority of whom they will go to travel with is friends; the second is going with family; and the third is club/or- ganization/foundation.

The researcher identifies the parameter of support at least 0.200. This value indicates the frequency of itemset at least 20% of all transaction. For instance, the itemsets that occurred 30% of all transaction are {reap and taste, learning lifestyle, transform} and {reap and taste, learning agriculture, learning lifestyle}. It means that the frequency of itemset {reap and taste, learning lifestyle, transform} occurs at least 30% of all transaction. The higher value of percentage represents the more interesting set of activities to be chosen.

According to Table 4, the activities that tourists would like to do when they visit orchard are fruit reaping and tasting (82.4%), walking in orchard (54.4%), learning agricultural tourism (52.2%), bicycling (52.2%), learning farmer's lifestyle (51.2%), respectively. For individual

activity, the highest support represents the most interesting activity which is fruit reaping and tasting.

Two activities with the highest support (50.7%) are fruit reaping and tasting and walking. For second order of support (47.5%), the two activities are fruit reaping and tasting and bicycling. With the third order of support (44.9%), the two activities are fruit reaping and tasting and learning agricultural tourism. With two activities, the highest interesting activities are fruit reaping and tasting including with walking.

For three activities: fruit reaping and tasting, walking, and learning agricultural tourism, have the highest support about 31.4%. The second rank of support for three activities: fruit reaping and tasting, walking, and shopping is 31.1%. The third order of support for three activities: fruit reaping and tasting, walking, and bicycling is 30.1%. With three activities, the highest interesting activities are fruit reaping and tasting, walking, and learning to do agriculture.

For four activities: fruit reaping and tasting, walking, bicycling and feeding animal, have the highest support about 21.8%. The second rank of support for four activities: fruit reaping and tasting, learning to do agriculture, learning lifestyle and learning to do product transformation is 21.6% which is the same as support value with these activities: fruit reaping and tasting, riding bicycle, learning lifestyle and learning to do product transformation. With four activities, the highest interesting activities are fruit reaping and tasting, walking, riding bicycle and feeding animal.

According to the result, fruit reaping and tasting activity is in every itemset no matter the size of activities is changed. It is implied that fruit reaping and tasting activity must be in every orchard.

However, only high support value is not enough to justify the correctness or accuracy to find out the association among activities even the criteria of support value is recommended to be high. However, the more increment of

Table 4
Frequent itemset with support above 0.200

Size	Support	Item 1	Item 2	Item 3	Item 4
1	0.824	Reap and taste			
1	0.544	Walking			
1	0.522	Learning agriculture			
1	0.522	Bicycling			
1	0.512	Learning lifestyle			
2	0.507	Reap and taste	Walking		
2	0.475	Reap and taste	Bicycling		
2	0.449	Reap and taste	Learning agriculture		
2	0.417	Reap and taste	Transform		
2	0.414	Reap and taste	Learning lifestyle		
2	0.412	Reap and taste	Shopping		
3	0.314	Reap and taste	Walking	Learning agriculture	
3	0.311	Reap and taste	Walking	Shopping	
3	0.301	Reap and taste	Walking	Bicycling	
3	0.300	Reap and taste	Learning lifestyle	Transform	
3	0.300	Reap and taste	Learning agriculture	Learning lifestyle	
4	0.218	Reap and taste	Walking	Bicycling	Feeding
4	0.216	Reap and taste	Learning agriculture	Learning lifestyle	Transform
4	0.216	Reap and taste	Bicycling	Learning lifestyle	Transform
4	0.208	Reap and taste	Walking	Learning agriculture	Shopping
4	0.208	Reap and taste	Walking	Bicycling	Shopping
4	0.208	Reap and taste	Walking	Shopping	Feeding

activities will influence the less value of support (North, 2012). As a consequence, for obtaining the better aspect of more recommended activities, the identified value of support should be recommended at least 0.200 which result shows the limitation in the number of activities not be more than 4 activities in an orchard.

With confidence more than 0.7, support more than 0.2 and lift more than 1 criteria, three to four activities are suggested to manage activities in the orchard.

According to Table 5, the association rules are identified;

1. Fruit reaping and tasting will be selected, given that waking, shopping, and feeding have been selected, with confidence about 100%.
2. Walking will be selected, given that fruit reaping and tasting, shopping, and feeding have been selected, with confidence about 86.7%.
3. Riding bicycle will be selected, given that fruit reaping and tasting, transform, and feeding have been selected, with confidence about 83.0%.
4. Learning farmer's lifestyle will be selected, given that learning agricultural tourism and learning agricultural product transformation have been selected, with confidence about 82.8%.
5. Learning agricultural product transformation will be selected, given that learning farmer's lifestyle and planting have been selected, with confidence about 81.2%.
6. Learning agricultural tourism will be selected, given that learning farmer's lifestyle and purchasing souvenir have been selected, with confidence about 77.8%.
7. Purchasing souvenir will be selected, given that fruit reaping and tasting, walking, and feeding have been selected, with confidence about 73.9%.
8. Feeding animal will be selected, given that riding bicycle and feeding animal have been selected, with confidence about 73.2%.

According to Table 6, the visitors are classified into 4 segments based on their interested activities. The first segment is visitors who are emphasized on only reaping and tasting the fruit. For the second segment, reaping and tasting the fruit is included. However, this segment is visitors who would like to learn how to do agriculture and lifestyle of agriculturists. Furthermore, the activities of the second segment are also walking in the orchard. The third segment is including reaping and tasting. The extended activities are feeding the animal and riding the bicycle. The

Table 6
Results of cluster analysis

Activities	Clusters			
	1	2	3	4
Fruit reaping and tasting	Yes	Yes	Yes	Yes
Walking in the orchard	No	Yes	No	Yes
Bicycling in the orchard	No	No	Yes	Yes
Learning to do agricultural	No	Yes	No	Yes
Attending festival	No	No	No	Yes
Feeding animal	No	No	Yes	Yes
Purchasing souvenir	No	No	No	Yes
Purchasing sapling or seed	No	No	No	Yes
Planting	No	No	No	Yes
Learning farmer's lifestyle	No	Yes	No	Yes
Learning agricultural product transformation	No	No	No	Yes
Business contact	No	No	No	No
Percentage of observation in each cluster	21%	30%	25%	24%

last segment is visitors who are interested all activities provided by the orchard except doing business contact.

The majority of observations is in cluster number 2; the visitors who are interested in reaping and tasting the fruit, learn how to do agriculture and lifestyle of agriculturists, and walking in the orchard. The second rank is in cluster number 3, 4, and 1, respectively. In addition, every clusters, the activity that all visitor is interested and must be in orchard is reaping and tasting the fruit.

Discussion and Recommendation

This research is studied for being the suggestion for arranging activities in orchard by using FP-Growth Algorithm. According to the support value, fruit reaping and tasting activity shows the highest support value. It implies this activity is important and highly suggested in orchard. Furthermore, the other activities that should be in orchard as following recommended are walking, shopping, and feeding because fruit reaping and tasting is influenced by these activities. Obviously, the pattern of activities is suitable for everyone in a big family.

The orchard owner should provide the instruments for reaping fruit to tourists and arrange the kiosk for tasting the fruit such providing buffet kiosk. Without these activities, the tourists are less interested to visit orchard. Furthermore, the pathway should be convenient and safe for walking inside orchard. The pathway should also have the architecture support for elder citizen because the travelers will be convenient to bring their family. Feeding animal is interesting activity for convincing children to visit

Table 5
Association rule of activities in orchard tourism (minimum confidence = 0.7 and lift = 1)

Rank	Premises	Conclusion	Support	Confidence	Lift
1	Walking, shopping, feeding	Reap and taste	0.208	1	1.214
2	Reap and taste, shopping, feeding	Walking	0.208	0.867	1.594
3	Reap and taste, transform, feeding	Bicycling	0.203	0.830	1.590
4	Learning agriculture, transform	Learning lifestyle	0.260	0.828	1.617
5	Learning lifestyle, planting	Transform	0.223	0.812	1.674
6	Learning lifestyle, shopping	Learning agriculture	0.206	0.778	1.490
7	Reap and taste, walking, feeding	Shopping	0.208	0.739	1.666
8	Bicycling, planting	Feeding	0.201	0.732	1.737

orchard. Shopping activity supports the housewife or teenager to purchase natural product from orchard as souvenir or personal use.

The strategic plan to convince travelers in cluster number 2 is providing the agricultural demonstration or display for them especially knowledge in philosophy of sufficient economy. They are interested to learn to do agriculture. For cluster number 3, providing bicycle rental is interested and also having the animal farm for motivating the visitors in this segment.

Conclusion

The data mining techniques are useful to evaluate and suggest the set of activities for orchard business owners to implement in order to convince visitors to travel. However, the evoked set of activities is the minimum requirement that business owners can consider. Development of orchard sightseeing should be continued and improved.

Conflict of interest

There is no conflict of interest.

References

Au, N., & Law, R. (2002). Categorical classification of tourism dining. *Annals of Tourism Research*, 29(3), 819–833.

Azevedo, P. J., & Jorge, A. M. (2007). Comparing rule measures for predictive association rules. Paper presented at the European Conference on Machine Learning.

Bach, M. P., Schatten, M., & Marusic, Z. (2013). Data mining applications in tourism: A keyword analysis. Paper presented at the Central European Conference on Information and Intelligent Systems.

Bose, I. (2009). Data mining in tourism. *Encyclopedia of Information science and technology* (2nd ed., pp. 936–940). Hershey, PA: IGI Global.

Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (1999). Using association rules for product assortment decisions: A case study. Paper presented at the Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: From concept to implementation*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Chajri, M., & Fakir, M. (2014). Application of data mining in e-commerce. *Journal of Information Technology Research (JITR)*, 7(4), 79–91.

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel & Tourism Marketing*, 12(1), 1–22.

García, E., Romero, C., Ventura, S., & Calders, T. (2007). *Drawbacks and solutions of applying association rule mining in learning management systems*. Paper presented at the Proceedings of the International Workshop on Applying Data Mining in e-Learning (ADML 2007), Crete, Greece.

Giraud-Carrier, C., & Povel, O. (2003). Characterising data mining software. *Intelligent Data Analysis*, 7(3), 181–192.

Girish, K. S., & Promila, S. (2016). A study on data mining algorithms for tourism industry. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, 7(1), 580–587. <http://dx.doi.org/10.21172/1.71.084>.

Karathiyra, M. B., Sakshi, R. S., Sakshi, D. S., & Kathiriya, D. R. (2012). Data mining for travels and tourism. *Journal of Information and Operations Management*, 3(1), 114.

Lantz, B. (2013). *Machine learning with R*. Birmingham, UK: Packt Publishing Ltd.

Lau, K.-N., Lee, K.-H., & Ho, Y. (2005). Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly*, 46(3), 344–362.

Liao, S.-H., Chen, Y.-J., & Deng, M.-Y. (2010). Mining customer knowledge for tourism new product development and customer relationship management. *Expert Systems with Applications*, 37(6), 4212–4223.

Merceron, A., & Yacef, K. (2008). *Interestingness measures for association rules in educational data*. Paper presented at the Educational Data Mining 2008.

Moscardo, G., Morrison, A. M., Pearce, P. L., Lang, C.-T., & O'Leary, J. T. (1996). Understanding vacation destination choice through travel motivation and activities. *Journal of Vacation Marketing*, 2(2), 109–122.

Mythili, M., & Shanavas, A. M. (2013). Performance evaluation of apriori and FP-growth algorithms. *International Journal of Computer Applications*, 79(10).

North, M. (2012). *Data mining for the masses*. Global Text Project.

Pesonen, J. (2013). Developing market segmentation in tourism: Insights from a Finnish Rural Tourism Study. *Publications of the University of Eastern Finland. Dissertations in Social Sciences and Business Studies*(69).

Petroman, I., & Petroman, C. (2010). *Agritourism and its forms*. Paper presented at the Scientific Conference “Durable agriculture-present and perspectives”, Iași.

Segall, R. S., & Zhang, Q. (2008). *Web mining of hotel customer survey data*. Paper presented at the Proceedings of the 12th Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2008), Orlando, FL, June.

Tama, B. A. (2015). Data mining for predicting customer satisfaction in fast-food restaurant. *Journal of Theoretical and Applied Information Technology*, 75(1).

Tan, P.-N. (2006). *Introduction to data mining*. New Delhi, India: Pearson Education India.

Tank, D. M. (2014). Improved apriori algorithm for mining association rules. *International Journal of Information Technology and Computer Science (IJITCS)*, 6(7), 15.

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), 2.