IMPLEMENTATION OF ASSOCIATION RULE METHOD IN BANGKA BELITUNG TOURISM APPLICATION IN ANDROID PLATFORM

Nio Somalo[1]

[1] Informatics Engineering – Indonesia Computer
Jl. Dipatiukur 112-114 Bandung
Email : niosomalo22@gmail.com[1]

ABSTRACT

Bangka Belitung province is an island province which consists of two major islands, the islands of Bangka and Belitung islands. Province holds the potential for tourism, such as marine tourism, and culinary history. However, little information about the tourism place make Bangka Belitung province not very well known by local and foreign tourists. Based on the interview and the difficulty of obtaining information on tourism place to be one of at least a factor of tourists coming to the province of Bangka Belitung.

One way to overcome these problems is to use data mining. Data mining techniques can be used to find recommendations for tourist-related activities and the association rule is a data mining techniques to find a correlation or an interesting pattern of a large data. One algorithm that can be used to find an interesting pattern is the a priori algorithm.
Construction of babel kite system using a priori algorithm predicted to produce recommendations related tourism place and activities. Based on the results of black box testing, beta and data mining is able to make a recommendation as tourism place.

Kata kunci*:* apriori, data mining, association rule, android

1 PENDAHULUAN

Information on tourism place is needed by tourists. By getting more information, to a tourist spot will make it easier for tourists to be able to make preparations and other needs. Based on interviews with tourists who have been to Bangka Belitung province, was found a few facts about the problems often experienced when visiting tourism place. A common issue is the tourists is difficult to get information about nearby tourims place, venue information, and information transport services available to visiting these sights. This is due to lack of information from the provincial offices of the Pacific Islands or from a travel agent.
Another phenomenon that occurred rating experiencing confusion in determining what activities can be done at a tourist spot or in some tourism place, because there are not the type of traveler looking for a place but the activity. Based on interviews rating stating that this confusion caused by yet available media that can help travelers to tell about activity what can be done on a travel destination, travelers normally only see from the habits of those who are in these sights, so travelers find it difficult to search for the information media and tourists usually only do by asking questions. Of the phenomenon takes on an activity or event.
Recommended activities and sights can be given to the traveler if there had been a history of other travelers and viewing patterns of linkage between the activities and tourism place that are often made by previous travelers. To be able to determine the right decision in recommending tours and activities to tourists, we need a method to be able to search for links between the tourism place and activities that are often carried by tourists that the association rule method.

1.1 Association Rule

Association rule is a procedure to look for relationships between items in a dataset [1]. Association rule (association rules) will find a certain pattern to associate the data with other data. To search for association rule from a data set, the first step that must be done is to find frequent of item set first. Frequent item set is a set of items that frequently appear simultaneously. After all frequent item set pattern is found, then look for a rule or rules associative linkages are eligible who have been determined [1].
If it is assumed that the goods are sold in supermarkets is the universe, then every item will have a boolean variable that will indicate the presence or not of goods in a single transaction or a shopping cart. Boolean patterns obtained are used to analyze items purchased simultaneously. The pattern formulated in an association rule. For example, consumers typically would buy computer and antivirus software at the same time shown as follows [1]:
Computer $\rightarrow$ antivirus software [*support* = 2, *confidence* = 60%]

Association rule required a variable size specified by the user to determine the limits of the extent to which or how much output the desired user. Support and confidence is a measure of confidence and usefulness of a pattern that has been found. Value 2 indicates that the overall support of the total transactions of consumers buying a computer and antivirus software at the same time as many as 2. While the 60% confidence that showed that if consumers buy computers and definitely buy the antivirus software by 60% [1].

1.1.1. Apriori Algorithm

Apriori algorithm is an algorithm that searches using the technique of frequent item set association rule. Algorithm uses apriori knowledge of the frequency of the attributes that have been previously known to process more information. On a priori algorithm, determines which candidates may appear in any way concerned minimum support and minimum confidence. Support is the percentage of the value of a visitor or a combination of an item in a database [1].

Values support an item obtained with the following formula: [1]

$$Support = (\text{A} \cap \text{B}) = \frac{\sum Transaction\ amunt\ which\ containing\ \text{A and B}}{\sum \text{Transaction total}} * 100\% \quad (2.1)$$

While confidence is certainty that the value of the strong relationship between items in a priori. Confidence can be sought after the emergence of an item frequency pattern found. The formula for calculating the confidence is as follows [1]:

$$Confidence = \frac{\sum Transaction\ amunt\ which\ containing\ \text{A and}}{\sum \text{Transaction which containing A}} * 100\% \quad (2.2)$$

The main process is carried out in a priori algorithm to obtain

frequent item set namely [1]:

1. Join (Merger)

This process is done by combining the item with other items that cannot be formed by the combination again.

2. Prune (trimming) The process of pruning that is a result of the items that have been combined and then trimmed with a minimum of support which has been determined by the user. At iteration k will be found all item set having k items, called the k-item set. Each iteration consists of two stages, namely:

a. Use frequent (k-1) item set to build candidate frequent k-item set.

b. Use database scan and pattern matching to collect counts for the candidate item sets.

The steps of the calculation process priori association rule algorithm as follows [1]:

1. The system scans the database to get the candidate 1-itemsets (the set of items consisting of 1 items) and calculate the value of its support. Then the value of its support is compared with a predetermined minimum support. If the value is greater than or equal to the minimum support, it includes large item set item sets.
2. Item sets that are not included in the large item set is not included in the next iteration (in crop).
3. In the second iteration, the system will use the results of large item sets in the first iteration (L1) to form a candidate item set second (L2). In the next iteration of the system will use the results of large item sets in the previous iteration (Lk-1) to form the following item set candidate (Lk). The system will combine (join) Lk-1 with Lk-1 to obtain Lk. As in previous iterations of the system will remove (cut) a combination of item sets that are not included in the large item set.
4. After a join operation, pairing new item set join the process results are calculated support her.
2. 5. The process of forming the candidate that consists of the merging process and cuts will continue until the set of candidate item sets to null, or is no longer a candidate to be formed.
3. 6. After that, the results of the frequent item set association rule established that meet the support and confidence that has been determined.
4. 7. On the formation of association rule, the same value is considered as one value.
5. 8. Association rule that is formed must meet specified minimum value.
6. 9. For every large item set L, find the subset L is not empty. For each of these subsets, generated rule to the form aB (L-a) if the support of its (a) greater than the minimum support.

1.2 Examination

System testing is done to find flaws or mistakes that exist in the system that is being tested. The test intends to determine the system that made it meets the criteria in accordance with the design objectives. There are three tests performed in this study, namely black box testing, beta testing, and acceptance testing.Pengujian *Black box*

Pengujian *black box* dilakukan dengan mengamati fungsional-fungsional yang terdapat dalam sistem dengan tujuan untuk mengetahui kesalahan pada sistem dan mengetahui apakah sistem telah sesuai dengan tujuan yang diharapkan.

1.2.1 Beta Testing

Beta testing is testing conducted objectively. tests are made directly to the respondents as users of the system to determine the extent of the feasibility of the system.

By doing beta testing, fault location as well as the shortcomings of the system is known to be improved in the coming stages of development. This test through two phases, namely a demo program, after the administration of the questionnaire.

1.2.2 Data Mining Testing

Data mining test plan to be carried out is by comparing the results on tourism place and activities that result from manual calculations using Microsoft Excel with the results of the software to be built. This test uses the whole data with a minimum amount of support 2 and a minimum value of 70% confidence.

2 RESEARCH

2.1 System Analysis

Systems analysis aims to identify the problems that arise in the construction of the system. The analysis includes problem analysis, analysis of similar applications, system architecture analysis, analysis of data sources, analysis of data preprocessing, application of methods of association rule analysis, analysis of software requirements specification, analysis of non-functional requirements, functional requirements analysis.

2.1.1 Problem Analysis

Here is the analysis of the problems arising from this study are:
1. The difficulty rating information facilities available
2. The difficulty rating to get information on tourism place and activities in tourist places by tourists.

2.1.2 System Overview

In general, the system built is a software to find patterns of relatedness between the sites to other tourism place and its activities. The parameters used in the analysis is the value of minimum support, minimum confidence value, user, time of transaction, tourism place, and activities.
The method used in the system to be built is a method to perform association rule frequent item set mining. To find frequent item set and determine the value of the support systems that would be built using a priori algorithm.

2.1.3 Data Analysis Selection

This research will be conducted two attributes that will be in mining, namely the attributes of tourism place and activities. Before the mining stage will be done first process to separate the attributes of the data selection of tourism place and activities. For the results of the selection process attribute sights can be seen in Table 1

Tabel 1 Data Selection

Timestamp	Name	Tourism Place
11/1/2015 15.59.59	Mochamad riezky nazrudin	Tanjung Kiras Beach, Air Anyer Beach, Batu Bedaun Beach
11/1/2015 16.14.54	Muhammad Iqbal	Merdeka Square Garden
11/1/2015 16.32.00	hendry wijaya	Tanjung Kalian Beach
11/1/2015 16.46.26	Rafiq zafir	Buku Limau Island, Dewi Kwan Yin Buddhist Monastery,
11/1/2015 17.19.05	I Gede Puja Arjana	Kampung Bali
11/1/2015 17.39.34	Humaira	Dewi Kwan Yin Buddhist Monastery, Buku Limau Beach
11/1/2015 17.50.45	Cepy Setia Nugraha	Hot Water Bathing Place Tirta Tapta Pemali
……	……	……

2.1.4 Cleaning Data Analysis

Data Cleaning is the process of removing noise. In this research, data cleaning process to eliminate where one transaction that has only one item will be eliminated. For data prior to the tourist spot cleaning process can be seen in Table 2. Following the results of the process of cleaning the data based on attributes of the tourism place.

Tabel 2 Data Cleaning

Timestamp	Name	Tourism Place
11/1/2015 15.59.59	Mochamad riezky nazrudin	Tanjung Kiras Beach, Air Anyer Beach, Batu Bedaun Beach
11/1/2015 16.46.26	Rafiq zafir	Buku Limau Island, Dewi Kwan Yin Buddhist Monastery,

Timestamp	Name	Tourism Place
11/1/2015 17.39.34	Humaira	Dewi Kwan Yin Buddhist Monastery, Buku Limau Beach,
11/1/2015 22.17.03	Agung Faishal Faris	Batu Bedaun Beach, Batu Belubang Beach, Mabai Beach
11/1/2015 23.02.10	bolly	Maria Cave, Batu Bedaun Beach, Bukit Batu Beach
11/2/2015 14.52.36	Silo	Tanjung Kalian Beach, Air Anyer Beach
11/2/2015 14.52.57	Asep Sarifudin	Romodong Beach, Air Anyer Beach
…….	……	…….

For activity data prior to the cleaning process can be seen in Table 1. Following the results of the process of cleaning the data based on the attributes of activity:

Timestamp	Name	Activity
11/5/2015 12.10.56	Bintang Januari Haliri	Taking pictures, eating, relaxing, swimming
11/5/2015 13.04.29	Agus Purwanto	Taking pictures, eating, relaxing, swimming
11/5/2015 13.56.13	Desy	Taking pictures, relaxing, swimming
11/5/2015 13.56.51	Sandro	Taking pictures, relaxing, swimming
11/5/2015 13.57.23	Husnaisa	Taking pictures, relaxing, swimming
11/14/2015 23.16.14	Gea	Taking pictures, Eating
11/14/2015 23.16.56	Cicilya	Relaxing, swimming
……...	……	…….

2.1.5 Data Mining Analysis

Application of association rule method in this study will be conducted by a travel and tourist activity to the minimum support used is two, and for a minimum of 70% confidence. As for the expectations in determining the two minimum support and minimum confidence of 70% is for the establishment of an efficient rule.

2.1.5.1 Tourism Place

The stages of the process algorithm is a priori in this study based on the attributes of the tourism place are as follows:

1. Generate 1-itemset Frequent Pattern

The first step is to scan the data of the sights in Table 2 to obtain candidate 1-itemset support count and calculate the value of each of the sights. Here are the results of the scan were performed:

Tabel 3 Kandidat 1-*itemset*

Tourism Place	Support Count
Tanjung Kiras Beach	2
Air Anyer Beach	4
Batu Bedaun Beach	6
Buku Limau Island	4
Dewi Kwan Yin Buddhist Monastery	3
Batu Belubang Beach	2
Mabai Beach	4
……	….

Remove (prune) the data that has the support count is less than a predetermined minimum support. Can be seen in Table 3 are 1-itemset candidate who has the support of less than the minimum support count, so it needs to be eliminated. Here are the results of candidates who meet the minimum support:

Tabel 4 Large 1-*itemset*

Tourism Place	Support Count
Tanjung Kiras Beach	2
Air Anyer Beach	5
Batu Bedaun Beach	6
Dewi Kwan Yin Buddhist Monastery	3
Batu Belubang Beach	2
Mabai Beach	4
Maria Cave	2
…….	…….

2. *Generate* 2-*itemset Frequent Pattern*

Furthermore, to find a pattern of frequent 2-itemsets or L2 combine (join) every item listed in Table 4, then

calculate its support count that will be obtained the following results:

Tabel 5 Candidate 2-*itemset*

Tourism Place	Support Count
Tanjung Kiras Beach, Air Anyer Beach	1
Tanjung Kiras Beach, Batu Bedaun Beach	2
Tanjung Kiras Beach, Dewi Kwan Yin Buddhist Monastery	0
Tanjung Kiras Beach, Batu Belubang Beach	0
Tanjung Kiras Beach, Mabai Beach	0
Tanjung Kiras Beach, Maria Cave	0
Tanjung Kiras Beach, Tanjung Kalian Beach	0
……	……

After the support count of every combination of 2-itemsets obtained, it is known that the combination of tourism place anywhere that does not meet the minimum support, so it must be removed (prune). Here are the results of a combination of 2-itemset meets the minimum support rules:

Tabel 6 Large 2-*itemset*

Tourism Place	Support Count
Tanjung Kiras Beach, Batu Bedaun Beach	2
Air Anyer Beach, Tanjung Kalian Beach	3
Bedaun Batu beach, Pantai Batu Belubang	2
Batu Bedaun Beach, Mabai Beach	2
Batu Bedaun Beach, Maria Cave	2
Batu Belubang Beach, Mabai Beach	2
Mabai Beach, Tanjung Pesona Beach	2

3. *Generate* 3-*itemset Frequent Pattern*
The next step is to combine (join) back of the sights in Table 6 to produce a candidate 3-itemsets, then calculated its support count, so the result will be obtained as follows:

Tabel 7 Candidate 3-*itemset*

Tourism Place	Support Count
Tanjung Kiras Beach, Batu Bedaun Beach, Air Anyer Beach	1
Tanjung Kiras Beach, Batu Bedaun Beach, Tanjung Kalian Beach	0
Tanjung Kiras Beach, Batu Bedaun Beach, Batu Belubang Beach	0
Tanjung Kiras Beach, Batu Bedaun Beach, Mabai Beach	0
Tanjung Kiras Beach, Batu Bedaun Beach, Maria Cave	0
Tanjung Kiras Beach, Batu Bedaun Beach, Tanjung Pesona Beach	0
Air Anyer Beach, Tanjung Kalian Beach, Batu Bedaun Beach	0
……	….

Remove (prune) the data that has the support count is less than a predetermined minimum support. Can be seen in Table 7 there are candidates who have a 3-itemset support count is less than the minimum support, so it needs to be eliminated. Here are the results of candidates who meet the minimum support:

Tabel 8 Large 3-*itemset*

Tourism Place	Support Count
Batu Bedaun beach, Batu Belubang Beach, Mabai Beach	2

The results of Table 8 shows that there is no combination of sights that can be done recombination, resulting item set search process stops. So frequent item set obtained as follows:

Tabel 9 Frequent itemset

Frequent Itemset
{Tanjung Kiras Beach}
{Air Anyer Beach}
{Batu Bedaun Beach}
{Dewi Kwan Yin Buddhist Monastery}
{Batu Belubang Beach}
{Mabai Beach}
{Maria Cave}
.....

4. Having obtained a frequent item set greater than or equal to the minimum support, the next step is to calculate the value of a frequent item set confidence that where the value of minimum confidence of 70%. Here are the results, where the calculation of the value

of the confidence that has resulted in a decimal value with the provision that if the value is greater after the decimal point is greater equal to five, then rounded up, while the number after the decimal value of less than five will be rounded bottom. For the calculation of the value of confidence and determine the pattern of association rule can be seen in table 10.

Tabel 10 Calculation of confidence values and determine patterns of association rule

Tourism Place	Confidence	Rule
If to {Batu Bedaun Beach} then to {Batu Belubang Beach and Mabai Beach}	2/6*100 %=33%	Rule Rejected
If to {Batu Belubang Beach} then to {Batu Bedaun Beach and Mabai Beach}	2/4*100 %=50%	Rule Rejected
If to {Mabai Beach} then to {Batu Bedaun Beach and Batu Belubang Beach}	2/4*100 %=50%	Rule Rejected
If to {Batu Belubang Beach and Mabai Beach} then to {Batu Bedaun Beach}	2/4*100 %=50%	Rule Rejected
If to {Batu Bedaun Beach and Mabai Beach} then to {Belubang Beach}	2/4*100 %=50%	Rule Rejected
If to {Batu Bedaun Beach and Batu Belubang Beach} then to {Mabai Beach}	2/4*100 %=50%	Rule Rejected
If to {Tanjung Kiras Beach} then to {Batu Bedaun Beach}	2/2*100 %=100 %	Rule Selected

2.1.6 Data Mining Test Result

Here are the results of apriori algorithm calculations manually compared with the results by babel kite system where the value of two minimum support and minimum confidence of 70% that has been incorporated in the system.

Tabel 11 Test Results data mining

Manual (*Microsoft excel*)	Babel Kite
if to Batu Belubang beach then to Batu Bedaun beach and Mabai beach	if to Maria cave then to Batu Bedaun beach
if to Tanjung Kiras beach then to Batu Bedaun beach	if to Batu Belubang beach then to Mabai beach
if to Batu Belubang beach then to Batu Bedaun beach	if to Tanjung Pesona then to Mabai beach
if to Maria cave then to Batu Bedaun beach	if to Tanjung Kiras beach then to Batu Bedaun beach
if to Batu Belubang beach then to Mabai beach	if to Batu Belubang beach then to Batu Bedaun beach and Mabai beach
if to Tanjung Pesona then to Mabai beach	if to Batu Belubang beach then to Batu Bedaun beach

3 CONCLUSION

Manual (*Microsoft excel*)	Babel Kite
if to Batu Belubang beach then to Batu Bedaun beach and Mabai beach	if to Maria cave then to Batu Bedaun beach
if to Tanjung Kiras beach then to Batu Bedaun beach	if to Belubang then to the beach Mabai
if to Batu Belubang beach then to Batu Bedaun beach	if to Tanjung Pesona beach then to Batu Bedaun beach
if to Maria cave then to Batu Bedaun beach	if to Tanjung Kiras then to Batu Bedaun beach
if to Batu Belubang beach then to Mabai beach	if to Batu Belubang beach then to Batu Bedaun beach and Mabai beach
if Tanjung Pesona beach then to Mabai beach	if to Batu Belubang beach then to Batu Bedaun beach

References

[1] Indonesia Travel. [Online]. HYPERLINK "http://www.indonesia.travel/destinations/destination/26/bangka-belitung"

http://www.indonesia.travel/destinations/destination/26/bangka-belitung

[2] Chandra. Babelprov. [Online]. HYPERLINK "http://babelprov.go.id/content/11298-wisatawan-kunjungi-babel" http://babelprov.go.id/content/11298-wisatawan-kunjungi-babel

[3] Micheline Kamber Jiawei Han, "Data Mining: Concepts and Technique," in *Data Mining: Concepts and Technique*. San Francisco: Diane Cerra, 2006, pp. 227-229.

[4] IDC Analyze the Future. [Online]. HYPERLINK "http://www.idc.com/prodserv/smartphone-os-market-share.jsp" http://www.idc.com/prodserv/smartphone-os-market-share.jsp

[5] MPA., Ph.D, Surya Dharma, *Pendidikan, Jenis, dan Metode Penelitian Pendidikan*. Jakarta: Direktorat Tenaga Kependidikan Department Pendidikan Nasional, 2008.

[6] Ian Sommerville, *Software Engineering*, Eight Edition ed.: Addison Wesley, 2007.

[7] Sommerville Ian, *Software Engineering*, Eight Edition ed.: Addison Wesley, 2007.

[8] KBBI. (2015, Desember) KBBI. [Online]. HYPERLINK "http://kbbi.web.id/" http://kbbi.web.id/

[9] Ida Nuraida, *Manajemen Administrasi Perkantoran*. Yogyakarta: Kanisius, 2008.

[10] Chr. Jimmy L. Gaol, *Sistem Informasi Manajemen*. Yogyakarta: Grasindo, 2008.

[11] Komaruddin Sastradipoera, *Asas-asas menejemen perkantoran*. Bandung: Kappa Sigma, 2001.

[12] gsbipb. (2013, Oct.) gsbipb. [Online]. HYPERLINK "http://gsbipb.com/?p=821" http://gsbipb.com/?p=821

[13] Nong Ye, *Data Mining: Theoris Algorithms and Examples*. NW, US: Taylor & Francis, 2014.

[14] Agus Wahadyo and Sudarma S, *Tip Trik Android untuk pengguna Tablet & Handphone*. Jakarta: Penerbit Mediakita, 2012.

[15] Putri Atalapu, "Implementasi Location Based Service Berbasis Cell Id Untuk Anjungan Provinsi Sulawesi Selatan Taman Mini Indonesia Indah (Tmii) Memanfaatkan Teknologi Augmented Reality Pada Perangkat Bergerak Android," p. 11, 2012.

[16] Harry J. Rosenblatt Gary Shelly, *Systems Analysis and Design*. Boston: Course technology cengange learning, 2010.

[17] Hanif Fatta, *Rekayasa Sistem Pengenalan Wajah*. Yogyakarta: Andi, 2009.

[18] Tanjung Hendri, Prabowo Haryi Marimin, *Sistem Infromasi Manajemen : Sumber Daya Manusia*. Bogor: Grasindo, 2006.

[19] Hendra Divayana, *Konsep OOAD*. Jakarta: STMIK Eresha, 2010.

[20] Julius Hermawan, *Analisa Desain & Pemrograman Berorientasi Objek dengan UML dan Visual Basic.NET*. Yogyakarta: Andi, 2010.

[21] Kim Hamilton and Russell Miles, *Learning UML 2.0*. United States of America: O'Reilly, 2006.

[22] Janner Simarmata, *Rekayasa Web*. Yogyakarta: ANDI, 2010.

[23] json org. json org. [Online]. HYPERLINK "http://json.org/json-id" http://json.org/json-id

[24] Marsela dan Veronica S. Moertini Yulita, *Analisis Keranjang Pasar dengan Algoritma Hash-Based pada transaksi penjualan di apotek*, 2002.

[25] S.Kom., M.Kom. Dicky Nofriansyah, *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta, Indonesia: Deepublish, 2015.

[26] Twinkle Puri, Binita Shah, Ishaan Bajaj, Binita Parekh Prof. Prashasti Kanikar, "Comparative Study of Apriori Algorithm Performance on Different Datasets," *Comparative Study of Apriori Algorithm Performance on Different Datasets*, vol. 4, no. 4, p. 40, April 2014.