Implementation of Apriori Algorithm in Determining Tourism Visit Patterns to Bali

Dewa Agung Krishna Arimbawa P[1*], I Nyoman Angga Prabawa[2], Putu Arya Mertasana[3]

[1.2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3]Department of Electrical and Computer Engineering, Udayana University
[*]dewagung.krishna@gmail.com

Abstract -Bali is a tourist destination terknal destinations internationally. Every month there are hundreds of thousands of tourists who visit Bali. With a high tourist visits, stakeholders must prepare all matters related to tourism in Bali. Because problems can occur when a visit in a particular month high but there is no readiness of the parties involved. The purpose of this study is to look for patterns of tourists visiting Bali using Apriori algorithm. Apriori algorithm is one method of Data Mining Based Rule-based Assossiation that can be used to identify patterns of events. Training data used is the visit of foreign tourists to Bali every month from 1982 to 2018. The calculation of Apriori algorithm also tested using an application that is Weka Data Mining. The results obtained are the rules of association which is a pattern of tourists visiting Bali. For example, "If tourists visiting Bali in September and November among the highest in the current year, the visit in November will not enter into the highest visit". By knowing the pattern of tourist arrivals in Bali, expected that the stakeholders can have a reference for decision making.

***Index of Terms*- Data Mining, Pattern Mining, Association, Rule Based, Apriori, Tourism.**

I. INTRODUCTION

Bali is a famous tourist destination destinations up to foreign countries. Thousands or even hundreds of thousands of foreign tourists visiting Bali every month. And the number of tourists visiting Bali each year is relatively increased. With the increasing tourist arrivals in Bali means that all relevant stakeholders have to prepare everything from infrastructure, human resources, accommodation, and so forth. But there are still problems-problems encountered when stakeholders can not predict about tourist arrivals to Bali which have an impact on the preparation of all relevant parties. For example, in a particular month tourist arrivals in Bali is booming but stakeholders are predicting this so there is no preparation, or it could be otherwise. For example, if stakeholders are not prepared if there pembludakan tourists visiting Bali then maybe this affects the image of Bali in the eyes of tourists. With the possible problems, we need a way to determine the pattern of tourist arrivals to Bali every month. How that can be done is to apply data mining to find patterns of tourists to visit Bali every month. There are many methods in data mining to choose one of them is a method of Apriori.

Data Mining is the term used to describe the extraction process value / information from a database [1]. To determine the pattern, can be used one method of data mining is the method of association. Association data mining is a job to determine the attributes to be obtained simultaneously. The task of the association is looking for rules that do not cover to measure the relationship between two or more attributes [2]. One well-known algorithms in data mining algorithms based association is a priori [3]. Priori algorithm aims to find a set of items most often in a set of data. Apriori algorithm defining a process to find a priori rule that meets the minimum requirements of the minimum requirements for the support and confidence [4]. In this research, priori calculation algorithms tested in an application Weka. Weka contains famous algorithms that Apriori algorithm for association rules data mining to find relationships between items within a given dataset [5].

Interest applied Apriori method is to be able to determine the pattern of tourist arrivals in Bali, so that it can serve as a reference for stakeholders to make decisions. In the end, better preparation will give a good image to Bali as a world tourist destination. Section II will be presented literature referenced in conducting research. Section III describes the methodology of research conducted. Section IV will present

the analysis and results of trials conducted. And last conclusion is described in Section V.

II. LITERATURE REVIEW

A. Data Mining

Data mining is the process of data analysis using software to find patterns and rules (rules) in the set of data. Data mining can analyze large amounts of data to find the knowledge to support decision making [6]. In addition to Data Mining is the process of extracting information from data collection through the use of algorithms and techniques that involve the field of statistics, machine learning, and database management system [7].

B. Association Rule

Analysis of the association or association rule mining is data mining techniques to discover the rules of the association between a combination of items [8]. Type association rule can be expressed as a percentage of the many activities that take place within a certain time [9]. In asisiasi analysis there are two components: Support and Confidence. Support is a measure that indicates the extent of the dominance of an item or itemset of the overall transaction. Confidence is a measure that
shows the relationship between the two items are conditional (based on a specific condition) [8].

C. apriori

Priori algorithm is data retrieval algorithms associative rules (Association rule) to determine the associative relationship of a combination of items. The a priori algorithm would be suitable to be applied when there are multiple relationships item you want analyzed. [7]. Alogritma priori properties Each subset frequent-itemset must be frequent-itemset [6].

D. Weka

Weka is software used for data mining tasks contain one Sattu famous algorithms that Apriori algorithm to perform data mining compute all the rules that have given minimum support and confidence given exceeds [5]. Weka give novice users a tool to find hidden information from the database and the file system with options and an easy to use visual interface [10].

III. METHODOLOGY

This research was conducted include the following steps.

1. Data source

The data used in this study came from the Badan Pusat Statistik (BPS) of Bali. The data used is the number of tourist arrivals to Bali every month from 1982 to 2018.

2. Document Review

Document Review conducted as the basis of knowledge and reference in the conduct of research. Document derived from related research journals.

3. Implementation of Apriori Algorithm

At this stage of implementation of phase-stages that exist in the Apriori algorithm to the data that had been prepared beforehand. Implementation is stated in the calculation according to the formulas and measures Apriori algorithm. At this stage also piloted the use of applications WEKA with Apriori method to the data that had been prepared.

4. Implementation Results

The result is a pattern of tourists visiting Bali rule based on the calculation of Apriori algorithm and application Weka.

5. withdrawal Conclusion

After getting the results of a pattern rule of Apriori algorithm calculations, then made a conclusion that can be used as a reference in the Bali tourism development,

,

IV. RESULT

A. Travel Visits Data

Here is the data highest tree tourist arrivals to Bali per month from 1982 to 2018 based on data Badap Pusat Statistik (BPS) of Bali.

TABLE I
TREE HIGHEST TOURISTS VISIT TO BALI
YEAR 1982 -2018

year	item Set
1982	Aug, July, Nov.
1983	Aug, Oct, Sept
1984	Aug, Mark, Jan.
1985	Aug, Des, Mark
1986	Aug, Des, Nov.
1987	Aug, Des, Oct
1988	July, Oct, Mark
1989	July, Sept, Des
1990	July, Aug, Sept
1991	July, Aug, Sept
1992	Aug, July, Sept
1993	Aug, July, Sept
1994	July, Aug, Sept
1995	Aug, July, Sept
1996	Aug, July, Sept
1997	Aug, Sept, July
1998	Aug, Sept, Oct
1999	Aug, July, Sept
2000	Aug, July, Sept
2001	Aug, July, Sept
2002	Aug, Sept, July
2003	Aug, July, Sept
2004	Aug, July, Sept
2005	Sep, July, Aug
2006	Des, July, Sept
2007	Aug, July, Sept
2008	Aug, July, Sept
2009	July, Aug, Des
2010	July, Aug, Sept
2011	July, Sept, Aug
2012	July, Des, Sept
2013	Aug, Sep, Oct
2014	July, Sept, Des
2015	Sep, July, Des
2016	July, Sept, Des
2017	Aug, July, Sept
2018	July, Aug, Sept

B. Tabular Data

Tabular data is transactional data that is the result of the processing of the raw data. Tabular Data containing the 3 highest ranked tourists to visit Bali each year. For example, in 1982 there were in August, July and November which became the highest ranked 3 tourists visiting Bali.

TABLE II
Tabular data TRANSACTIONS

year	Aug	Jul	Nov	Oct	Sep	Mar	Jan	Dec
1982	Y	Y	Y	N	N	N	N	N
1983	Y	N	N	Y	Y	N	N	N
1984	Y	N	N	N	N	Y	Y	N
1985	Y	N	N	N	N	Y	N	Y
1986	Y	N	Y	N	N	N	N	Y
1987	Y	N	N	Y	N	N	N	Y
1988	N	Y	N	Y	N	Y	N	N
1989	N	Y	N	N	Y	N	N	Y
1990	Y	Y	N	N	Y	N	N	N
1991	Y	Y	N	N	Y	N	N	N
1992	Y	Y	N	N	Y	N	N	N
1993	Y	Y	N	N	Y	N	N	N
1994	Y	Y	N	N	Y	N	N	N
1995	Y	Y	N	N	Y	N	N	N
1996	Y	Y	N	N	Y	N	N	N
1997	Y	Y	N	N	Y	N	N	N
1998	Y	N	N	Y	Y	N	N	N
1999	Y	Y	N	N	Y	N	N	N
2000	Y	Y	N	N	Y	N	N	N
2001	Y	Y	N	N	Y	N	N	N
2002	Y	Y	N	N	Y	N	N	N
2003	Y	Y	N	N	Y	N	N	N
2004	Y	Y	N	N	Y	N	N	N
2005	Y	Y	N	N	Y	N	N	N
2006	N	Y	N	N	Y	N	N	Y
2007	Y	Y	N	N	Y	N	N	N
2008	Y	Y	N	N	Y	N	N	N
2009	Y	Y	N	N	N	N	N	Y
2010	Y	Y	N	N	Y	N	N	N
2011	Y	Y	N	N	Y	N	N	N
2012	N	Y	N	N	Y	N	N	Y
2013	Y	N	Y	N	Y	N	N	N
2014	N	Y	N	N	Y	N	N	Y
2015	N	Y	N	N	Y	N	N	Y
2016	N	Y	N	N	Y	N	N	Y
2017	Y	Y	N	N	Y	N	N	N
2018	Y	Y	N	N	Y	N	N	N

C. Analia High Frequency Pattern (Support)

This stage is looking for a combination of items that meet the minimum requirements of the value of the support in the data.

1. Formation 1 itemsets

The process of formation of C1 with 1 itemsets. Here's one example of the calculation 1 itemsets:

$$Support\ (aug) = \frac{Aug\ Transaction}{Total\ Transaction}\ x\ 100\%$$
$$Support\ (aug) = \frac{30}{37}\ x\ 100\%$$
$$Support\ (aug) = 81.08\%$$

and so on until December Support

TABLE III
SUPPORT 1 ITEM

	Total	Support
Aug	30	81.08%
July	30	81.08%
November	3	8:11%
Oct	4	10.81%
boss	30	81.08%
March	3	8:11%
January	1	2.70%
dec	10	27.03%

2. Formation of 2 itemset

The process of formation of C2 with 2 itemset. Here's one example of calculation 2 itemset:

$$Support\ (aug, jul) = \frac{Aug\ Jul\ Transaction}{Total\ Transaction}\ x\ 100\%$$
$$Support\ (aug, jul) = \frac{23}{37}\ x\ 100\%$$
$$Support\ (aug, jul) = 62.16\%$$

and so on until the Support of Jan, Dec.

TABLE IV
SUPPORT 2 ITEMS

	Total	Support
Aug & July	23	62.16%
Aug & Nov	3	8:11%
Aug & Oct	3	8:11%
Aug & Sep	24	64.86%
Aug & Mar	2	5:41%
Aug & Jan	1	2.70%
Aug & Sep	4	10.81%
Jul & Nov	1	2.70%
Jul & Oct	1	2.70%
Jul & Sep	27	72.97%
Jul & Mar	1	2.70%
Jul & Jan	0	0.00%
Jul & Dec	7	18.92%
Nov & Oct.	0	0.00%
Nov & Sep	1	2.70%
Nov & Mar	0	0.00%
Nov & Jan	0	0.00%
Nov & Des	1	2.70%
Oct & Sep	2	5:41%
Oct & Mar	1	2.70%
Oct & Jan	0	0.00%
Oct & Des	1	2.70%
Sep & Mar	0	0.00%
Sep & Jan	0	0.00%
Sept & Des	6	16:22%
Mar & Jan	1	2.70%
Mar & Dec	1	2.70%
Jan & Dec	0	0.00%

The calculation of the amount passed on to 4 itemset itemsets.

D.Formation of Associative Rules (Confidence)

After all the high frequency pattern is found, then sought associative rules that meet the minimum requirements for confidence by calculating confidence associative rule A → B. one example of the confidence value of the rule A → B is obtained as follows.

$$Confidence\ (aug, jul) = \frac{Aug\ Jul\ Transaction}{Aug\ Transaction}\ x\ 100\%$$

$$Support\ (aug, jul) = \frac{23}{30}\ x\ 100\%$$

$$Support\ (aug, jul) = 76.67\%$$

and so on until the Confidence of Jan, Dec.

TABLE V
CONFIDENCE 2 ITEMS

	total	confidence
Aug & July	23	76.67%
Aug & Nov	3	10:00%
Aug & Oct	3	10:00%
Aug & Sep	24	80.00%
Aug & Mar	2	6.67%
Aug & Jan	1	3:33%
Aug & Sep	4	13:33%
Jul & Nov	1	3:33%
Jul & Oct	1	3:33%
Jul & Sep	27	90.00%
Jul & Mar	1	3:33%
Jul & Jan	0	0.00%
Jul & Dec	7	23:33%
Nov & Oct.	0	0.00%
Nov & Sep	1	33.33%
Nov & Mar	0	0.00%
Nov & Jan	0	0.00%
Nov & Des	1	33.33%
Oct & Sep	2	50.00%
Oct & Mar	1	25.00%
Oct & Jan	0	0.00%
Oct & Des	1	25.00%
Sep & Mar	0	0.00%
Sep & Jan	0	0.00%
Sept & Des	6	20.00%
Mar & Jan	1	33.33%
Mar & Dec	1	33.33%
Jan & Dec	0	0.00%

Fig. 1 shows a comparison of the values of support and confidence in 2 set items using a graph.

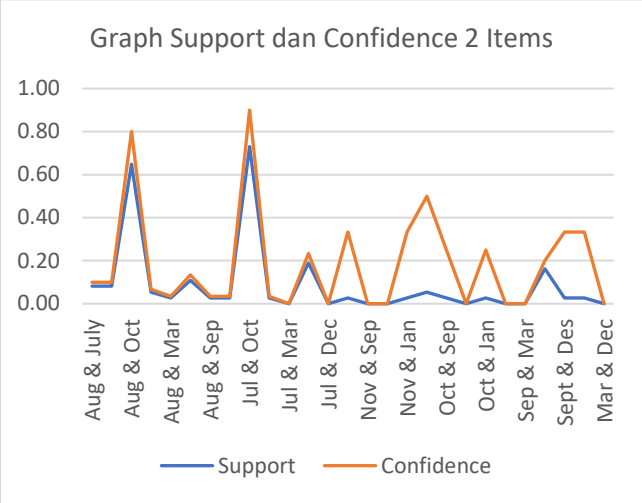

Fig 1. Graph Support and Confidence 2 items

E. Results of Applications WEKA

Weka are used to perform mining based on data prepared by the user. So users do not need to do manual counting or even make a special application. Weka has a library to do the calculation algorithm Apriori. Here is the result of the data application tests Weka tourists to visit Bali every month from 1982 to 2018. The minimum value of support is 15% and minimum is 60% Confidence.

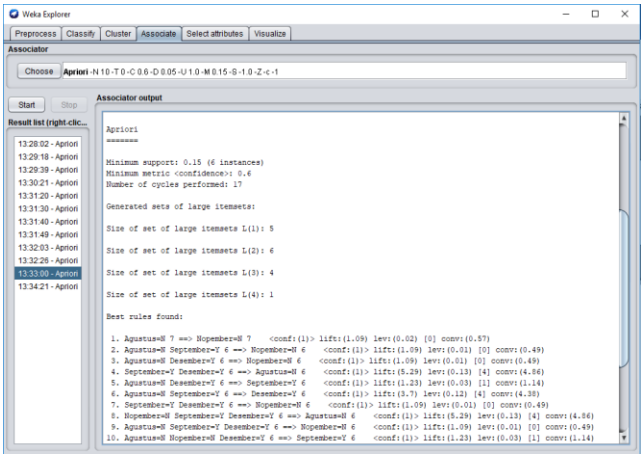

Fig 2. The calculation results with the application Weka

```
Apriori
=======

Minimum support: 0.15 (6 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1
```

Fig 3. Setting the support, confidence Applications WEKA

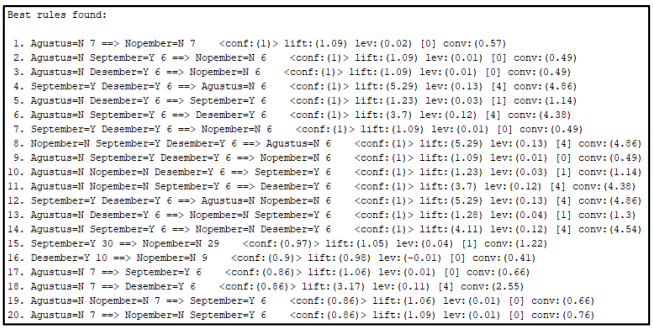

Fig 4. Results Best Rule of application Weka

Based on calculations using the application Weka Apriori association rules then obtained best with confidence values respectively. Confidence that the value 1 means that the rules must have happened during 1982 to 2018. For example Rule 7 which states that "If the tourists visiting Bali in September and November among the highest in the current year, the visit in November will not enter into the highest visit ".

V. CONCLUSION

Bali as a world tourism destination has to be prepared with tourist arrivals increasing. By knowing the pattern of tourist arrivals in Bali, all relevant stakeholders can prepare everything so that it can maintain the quality of Bali travel to tourists. Pattern mining is done using the Apriori algorithm which is one method of data mining based association rule-based. Training data is the number of tourist arrivals to Bali each month from 1982 to 2018. The training data comes from the Badan Pusat Statistik (BPS) of Bali. Apriori algorithm implementation process carried out in stages starting from the calculation Support, Confidence, and rules. Weka application also implemented to perform calculations quickly to training data. The results obtained are the rules which is a pattern of tourists visiting Bali. For example, "If tourists visiting Bali in September and November among the highest in the current year, the visit in November will not enter into the highest visit". By knowing the pattern of tourist arrivals in Bali, is expected to be a reference for stakeholders to take the policy in order to maintain the image of Bali in the eyes of world travel.

REFERENCES

[1] Hendrayana Putu Bagus Surya, Rifky Lana Rahardian, and Made Sudarma. “Application of Neural Network Overview in Data Mining”, International Journal of Engineering and Emerging Technology, Vol. 2, No. 1, January—June 2017.

[2] Maharani et al. “Implementasi Data Mining Untuk Pengaturan Layout Minimarket Dengan Menerapkan Association Rule”. Jurnal Riset Komputer (JURIKOM), Vol. 4 No. 4, Agustus 2017.

[3] Adi Panca Saputra Iskandar, Kheri Arionadi Shobirin, and Komang Oka Saputra. “Analysis of Shopping Cart at Drugs Store by Using an Apriori Algorithm”. International Journal of Engineering and Emerging Technology, Vol. 2, No. 1, January—June 2017.

[4] Putu Bagus Indra Sukadiana Putra, Ni Putu Sri Merta Suryani, and Sri Aryani. “Analysis of Apriori Algorithm on Sales Transactions to Arrange Placement of Goods on Minimarket”. International Journal of Engineering and Emerging Technology, Vol. 3, No. 1, January—June 2018.

[5] HarvinderChauhan, AnuChauhan. “Implementation of the Apriori algorithm for association rule mining”. COMPUSOFT, An international journal of advanced computer technology, Volume-III, Issue-IV, April-2014.

[6] Dana Sulistiyo Kusumo, Moch. Arief Bijaksana, Dhinta Darmantoro. “Data Mining Dengan Algoritma Apriori Pada Rdbms Oracle”. Jurnal Penelitian dan Pengembangan TELEKOMUNIKASI, Vol. 8 No. 1, Juni 2003.

[7] Robi Yanto, Riri Khoiriah. “Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat”. Citec Journal, Vol. 2, No. 2, Februari 2015 – April 2015.

[8] Goldie Gunadi, Dana Indra Sensuse. “Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (FP-Growth): Studi Kasus Percetakan PT. Gramedia”. Jurnal TELEMATIKA MKOM, Vol.4 No.1, Maret 2012.

[9] Paska Marto Hasugian. “Pengujian Algoritma Apriori Dengan Aplikasi Weka Dalam Pembentukan Asosiation Rule”. Jurnal Mantik Penusa, Volume 1, No. 2, Desember 2017.

[10] Faisal Mohammed Nafie Ali & Abdelmoneim Ali Mohamed Hamed. “Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents”. Journal of Information and Telecommunication, VOL. 2, NO. 3, 231–245, 2018.