Souvenir Sales Analysis using Apriori Algorithm (Case Study: Ubud-Market Transaction in March 2020)

Muhammad Firyanul Rizky[a1], Ida Bagus Made Mahendra[a2], I Gusti Agung Gede Arya Kadyanan[a3]

[a]Informatics Departement, Faculty of Math and Science, Udayana University
Bali, Indonesia
[1]firyanulrizky@student.unud.ac.id
[2]ibm.mahendra@unud.ac.id
[3]gungde@unud.ac.id

Abstract

Ubud market is one of the largest art markets in Bali, there are many local Balinese souvenir traders and craftspeople, most of them are livelihoods depend on buying and selling local souvenirs, Since the Covid-19 pandemic entered in April 2020, Ubud market traders have started to close their business and hoping economic recoveryin future. The author tries to do a track record of souvenir sales transactions in Ubud market to find the last sales pattern before the traders closes their business to give a solution for marketing strategies in future. The sales transaction data will just become meaningless trash if it's useless.. To get use information about the products that are most sold out at Ubud Market from the transaction database, the author uses the Apriori algorithm. This study was determined final rules on 2 itemset combination, If buying Manik-Manik Craft, Also buy Barong Shirt with the highest confidence 70% and Minimum Support 28%, and for 3 itemset a combination, If buying Celuk Silver, and Barong Shirt, Also buy Manik-Manik Craft with the highest confidence 37.5% and Minimum Support 12%, based on that there are 3 best-selling souvenir products, namely Barong Shirt, Manik-Manik Craft and Silver-Celuk in March 2020.

Keywords: Apriori Algorithm, Data Mining, Sales Analysis, Association Rule Mining, Ubud Market.

1.   Introduction

Ubud art market is one of the largest art markets in Bali, there are many local Balinese souvenir traders and craftspeople, most of them are livelihoods depend on buying and selling local souvenirs, in daily sale transactions, Since the Covid-19 pandemic entered in April 2020, Ubud market traders have started to close their business and can only hope in economic recovery. The author tries to do a track record of souvenir sales transactions in March 2020 to find the last sales pattern before the traders closes their business to give a solution for marketing strategies in future. The method for getting data transaction is by Interviewing and asking for transaction notes from several traders who owned houses and businesses in the Ubud market area as well as some tourists who were still passing by at that location.

The apriori algorithm is the most well-known algorithm for finding high frequency patterns. Apriori is divided into several stages called narration or pass. To form candidate itemsets, candidate k-itemsets are formed from the combination (k-1) -itemsets obtained from the previous iteration. One way of the a priori algorithm is the trimming of k-itemset candidates whose subsets containing k-1 items are not included in the high frequency pattern with length k-1. Apriori algorithm is well known for finding high frequency patterns. The process of forming a combination of itemsets and creating rules begins with data analysis.

The data used are the sales transaction data for souvenir in Ubud Market, the data was analyzed in March 2020 to find out which eyewear products are most sold out in one month, then proceed with the formation of the itemsets combination pattern and from the interesting itemsets combination pattern, association rules are formed. In this case, the Apriori algorithm is

very suitable for the case of sales data processing that wants to find out which products are the most sold out.

This research taking references from previous studies: the application of apriori algorithm to find patterns of association rules in a temporal data in the database[1]. This study is adding analysis for 3 candidates itemset generation to get build 3 combinations of association rules.

2. Reseach Methods

In this research, the author used apriori algorithm aims to find all apriori rules that meet the minimum requirements of support (Support value), namely the combination of each item in the database. And the minimum confidence requirement (certainty value), namely the strength of the relationship between items in the association rule based on souvenir sales transactions at Ubud Market.

2.1. Apriori Algorithm

The apriori algorithm is the best known algorithm for finding high frequency patterns. the apriori algorithm is divided into several stages called narration or pass[2]. Here is the basic concept of the apriori algorithm :

a. The formation of the itemset candidate, the k-itemset candidate is formed from the (k-1) -itemset combination obtained from the previous iteration. One way of the apriori algorithm is the trimming of k-itemset candidates whose subsets containing k-1 items are not included in the high frequency pattern with length k-1.

b. Support calculation for each k-itemset candidate. Support from each k-itemset candidate is obtained by scanning the database to calculate the number of transactions that contain all the items in the k-itemset candidate. This is also a feature of the a priori algorithm where it is necessary to calculate the entire database as many as the longest k-itemset.

c. Set high frequency pattern. High frequency patterns containing k items or k-itemset are determined from k-itemset candidates whose support is greater than the minimum support. If no new high frequency patterns are found then the whole process is terminated. If not, then k is added by one and returns part 1.

Process stages in the Apriori Algorithm:

a. High Frequency Pattern Analysis with Apriori Algorithm This stage looks for item combinations that meet the minimum requirements of the support value in the database. The support value of an item is obtained using the following formula :

$$Support\ A = \frac{\sum Transactions\ containing\ A}{\sum Transaction} \times 100\%$$

(1)

b. The Temporary 2-Itemset until k-itemset Combination, the support value of 2 until k-items is obtained using the formula: Support (A, B) = P (A∩B)

$$Support\ A, B = \frac{\sum Transactions\ containing\ A\ and\ B}{\sum Transaction} \times 100\%$$

(2)

the occurrence is more than the specified minimum value (Ø). Suppose Ø = 2, then all itemsets whose occurrence frequency is more than or equal to 2 times are called frequent. The set of frequent k itemset is denoted by Fk.

c. Formation of Association Rules After all high frequency patterns found, then look for the association rule that meets the minimum requirements for confidence by calculating the associative confidence rule A-> B. The confidence value of the rule A-> B is obtained by the following formula:

$$\text{Confidence} = \frac{\sum \text{Transactions containing A and B}}{\sum \text{Transaction } A} \times 100\%$$

(3)

To determine the association rules to be selected, it must be sorted based on Support ×
Confidence. N rules are taken that have the greatest payoff.

3. Result and Discussion

3.1. Problem Analysis at Ubud Market

The transaction data sales on Ubud Market so far are not well structured, so the sales data
which is getting more and more only serves as an archive for the company and cannot be used
by the shop for developing marketing strategies. Therefore, the Ubud Market trader needs a
system to process data that can produce sales data for the most frequently purchased souvenir
products, so that the most frequently purchased souvenir products can be used as a reference
for developing a marketing strategy for these products to consumers.

3.2. List of Souvenir Products at Ubud Market

This list is some samples of famous souvenirs and the most commonly found and sold in the
Ubud market, this is validated from the author's interviews with several traders and tourists who
made transactions in March 2020, this list can be seen in this table below:

Table 1. Top 15 Souvenir Product List in Ubud Market

Item Code	Product Name	Type
01	Joger Shirt	Fashion
02	Balinese Batik	Fashion
03	Manik-Manik Craft	Handycraft
04	Barong Shirt	Fashion
05	Mukena Bali	Fashion
06	Bali Home Spa	Aromatherapy
07	Perak Celuk	Accessories
08	Ethnic Sandals	Fashion
09	Woven Cloth Endek	Fashion
10	Traditional Painting	Art
11	Layang-Layang	Handycraft
12	Kebaya	Fashion
13	Palm Leaf Painting	Art
14	Dreamchatcher	Handycraft
15	Pandora Box	Handycraft

3.3. Transaction Patterns of Souvenir Product Sales

Based on sales transactions of souvenir products at Ubud Market, these transactions can be
accumulated by top 10 transaction. The accumulated sales transactions of souvenir products at
the Ubud Market are obtained from daily sales, the transaction data can be seen in this table
below :

Table 2. Top 10 Souvenir Transaction List in Ubud Market in March 2020

KDTrans	Products Sold Out	Transaction Date
01	Joger Shirt , Barong Shirt , Woven Cloth Endek ,	2020-03-01
02	Joger Shirt , Balinese Batik , Barong Shirt ,	2020-03-01
03	Manik-Manik Craft , Barong Shirt , Woven Cloth Endek ,	2020-03-03

04	Joger Shirt , Balinese Batik , Barong Shirt ,	2020-03-04
05	Joger Shirt , Barong Shirt , Woven Cloth Endek ,	2020-03-05
06	Joger Shirt , Manik-Manik Craft , Woven Cloth Endek ,	2020-03-08
07	Joger Shirt , Balinese Batik , Woven Cloth Endek ,	2020-03-07
08	Joger Shirt , Balinese Batik , Manik-Manik Craft ,	2020-03-08
09	Balinese Batik , Manik-Manik Craft , Barong Shirt ,	2020-03-08
10	Joger Shirt , Manik-Manik Craft , Barong Shirt ,	2020-03-09

3.4. Itemset Generation (1 Itemset)

The following is a solution with a case example based on the data provided in table 2. The process of generating candidate 1 itemset or we abreviation it with C1, this C1 is containing a minimum amount of support = 40% with the following formula :

$$\text{Support } A = \frac{\sum \text{Transactions containing } A}{\sum \text{Transaction}} \times 100\%$$

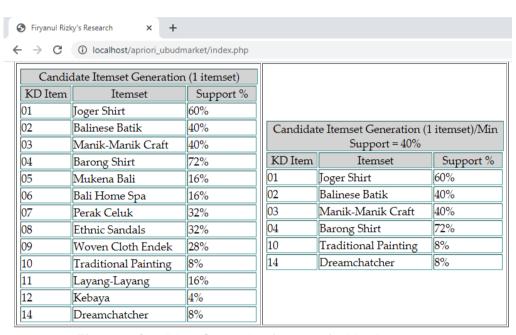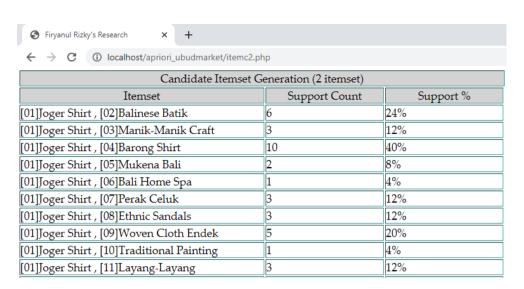(4)

Figure 1. Candidate Generation (1 Itemset) with minsup 40%

3.5. Candidat Generation (2 Itemset)

The process of generating candidate 2 itemset or we abreviation it with C2, it can be solved with the following Support (A, B) = P (A∩B) formula below :

$$\text{Support } A, B = \frac{\sum \text{Transactions containing } A \text{ and } B}{\sum \text{Transaction}} \times 100\%$$

(5)

Candidate Itemset Generation (2 itemset)		
Itemset	Support Count	Support %
[01]Joger Shirt , [02]Balinese Batik	6	24%
[01]Joger Shirt , [03]Manik-Manik Craft	3	12%
[01]Joger Shirt , [04]Barong Shirt	10	40%
[01]Joger Shirt , [05]Mukena Bali	2	8%
[01]Joger Shirt , [06]Bali Home Spa	1	4%
[01]Joger Shirt , [07]Perak Celuk	3	12%
[01]Joger Shirt , [08]Ethnic Sandals	3	12%
[01]Joger Shirt , [09]Woven Cloth Endek	5	20%
[01]Joger Shirt , [10]Traditional Painting	1	4%
[01]Joger Shirt , [11]Layang-Layang	3	12%

Figure 2. Top 10 Candidate Generation (Itemset 2)

In this case, the author uses the minimum support specified is 25% to get best candidate generation in 2 itemset, so the combination of 2 itemsets that do not meet the minimum support will be pruning, then we get just 2 candidate, as shown in the table below :
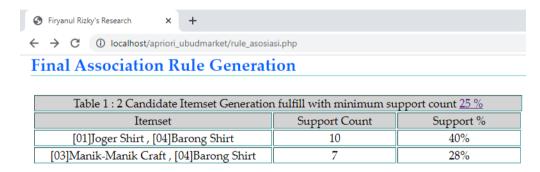
Final Association Rule Generation

Table 1 : 2 Candidate Itemset Generation fulfill with minimum support count 25 %		
Itemset	Support Count	Support %
[01]Joger Shirt , [04]Barong Shirt	10	40%
[03]Manik-Manik Craft , [04]Barong Shirt	7	28%

Figure 3. Best Candidate Generation (Itemset 2) with minsup 25%

3.6. 3 Candidat Itemset Generation

The process of forming C3 or called 3 itemset with a minimum amount of support = 10% Can be completed with the following formula : Support (A, B) = P (A∩B∩C)

$$\text{Support } A, B, C = \frac{\sum \text{Transactions containing A, B and C}}{\sum \text{Transaction}} \times 100\%$$

(6)

Candidate Itemset Generation (3 itemset)		
Itemset	Support Count	Support %
[01]Joger Shirt , [02]Balinese Batik , [03]Manik-Manik Craft	1	4
[01]Joger Shirt , [02]Balinese Batik , [14]Dreamchatcher	0	0
[01]Joger Shirt , [02]Balinese Batik , [12]Kebaya	0	0
[01]Joger Shirt , [02]Balinese Batik , [11]Layang-Layang	1	4
[01]Joger Shirt , [02]Balinese Batik , [10]Traditional Painting	1	4
[01]Joger Shirt , [02]Balinese Batik , [09]Woven Cloth Endek	1	4
[01]Joger Shirt , [02]Balinese Batik , [08]Ethnic Sandals	0	0
[01]Joger Shirt , [02]Balinese Batik , [07]Perak Celuk	0	0
[01]Joger Shirt , [02]Balinese Batik , [06]Bali Home Spa	0	0
[01]Joger Shirt , [02]Balinese Batik , [05]Mukena Bali	0	0
[01]Joger Shirt , [02]Balinese Batik , [04]Barong Shirt	3	12

Figure 4. Top 10 Candidate Generation (Itemset 3)

In this case, the author uses the minimum support specified is 12% to get minimal one of the best candidate generation in 3 itemset, so the combination of 3 itemsets that do not meet the minimum support will be pruning, then we get just 3 candidate, as shown in this table below :

Table 2 : 3 Candidate Itemset Generation fulfill with minimum support count 12 %		
Itemset	Support Count	Support %
[01]Joger Shirt , [02]Balinese Batik , [04]Barong Shirt	3	12%
[03]Manik-Manik Craft , [04]Barong Shirt , [07]Perak Celuk	3	12%

Figure 5. Best Candidate Generation (Itemset 3) with minsup 12%

3.7. Association Rules Extraction

After all high frequency patterns are found, then look for the association rule that meets the minimum requirements for confidence by calculating the associative confidence rule A → B. Confidence value from rule A → B is obtained by the following formula:

$$\text{Confidence} = \frac{\sum \text{Transactions containing A and B}}{\sum \text{Transaction } A} \times 100\%$$

(7)

From the combination of the 2 itemset that has been found, it can be seen the amount of support value and confidence of the candidate association rules as shown in the table below :

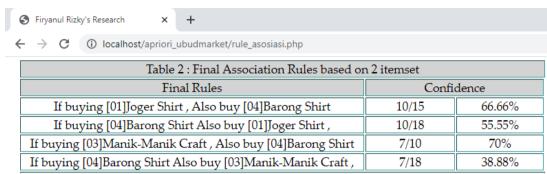| Table 2 : Final Association Rules based on 2 itemset		
Final Rules	Confidence	
If buying [01]Joger Shirt , Also buy [04]Barong Shirt	10/15	66.66%
If buying [04]Barong Shirt Also buy [01]Joger Shirt ,	10/18	55.55%
If buying [03]Manik-Manik Craft , Also buy [04]Barong Shirt	7/10	70%
If buying [04]Barong Shirt Also buy [03]Manik-Manik Craft ,	7/18	38.88%

Figure 6. Final Association Rules based on 2 itemset

Then also From the combination of the 3 itemset that has been found one rules that can be seen the amount of support value and confidence of the candidate association rules as shown in the table :
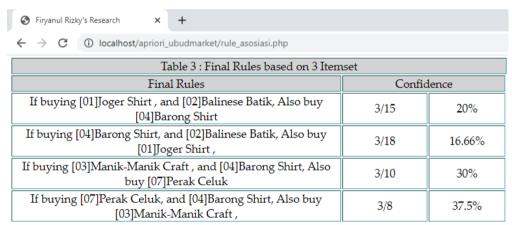
Table 3 : Final Rules based on 3 Itemset		
Final Rules	Confidence	
If buying [01]Joger Shirt , and [02]Balinese Batik, Also buy [04]Barong Shirt	3/15	20%
If buying [04]Barong Shirt, and [02]Balinese Batik, Also buy [01]Joger Shirt ,	3/18	16.66%
If buying [03]Manik-Manik Craft , and [04]Barong Shirt, Also buy [07]Perak Celuk	3/10	30%
If buying [07]Perak Celuk, and [04]Barong Shirt, Also buy [03]Manik-Manik Craft ,	3/8	37.5%

Figure 7. Final Association Rules based on 3 itemset

3.8. Final Association Rules

The final association rules are ordered based on the largest confidence that has been determined, can be seen in the table below :

Table 3. Final Association Rules ordered by largest confidence

Final Rules	Confidence	
If buying Manik-Manik Craft , Also buy Barong Shirt	7/10	70%
If buying Joger Shirt , Also buy Barong Shirt	10/15	66,66%
If buying Perak Celuk, and Barong Shirt, Also buy Manik-Manik Craft	3/8	37,5%
If buying Joger Shirt , and Balinese Batik, Also buy Barong Shirt	3/15	20%

Based on the association rules above, it can be seen that the most sold out souvenir products in March 2020 at the Ubud Market that can be seen in the following graph below :
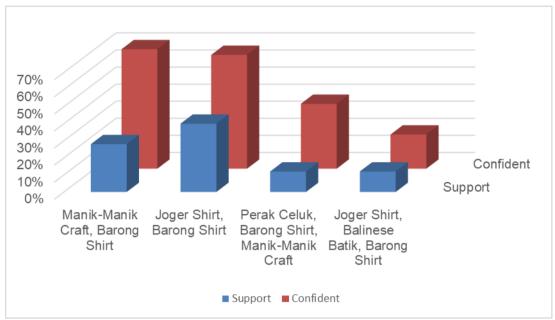
Figure 9. Chart Result of Association Rules Visualisation

4. Conclusion

The conclusion of this research is based on the graphic above determined final rules on 2 itemset combination, If buying Manik-Manik Craft, Also buy Barong Shirt with the highest confidence 70% and Minimum Support 28%, and for 3 itemset a combination, If buying Celuk Silver, and Barong Shirt, Also buy Manik-Manik Craft with the highest confidence 37.5% and Minimum Support 12%, based on that there are 3 best-selling souvenir products, namely Barong Shirt, Manik-Manik Craft and Silver-Celuk in March 2020, with the knowledge of the most sold out products, so that Ubud Market trader can prepare to develop a marketing strategy after pandemic Covid-19 to making other products by examining what the most advantages of the product sold with other products and can increase the supply of Manik-Manik Craft, Barong Shirt, Joger Shirt and Perak Celuk in post pandemic.

1. Sales of the most sold souvenir products at Ubud Market can be determined using the apriori algorithm, by looking at the products that meet the minimum support and minimum confidence, the most sold products are Manik-Manik Craft, Barong Shirt, Joger Shirt and Perak Celuk, but in calculating support and confidence difficult if the data is processed in large amounts.
2. The Apriori algorithm can help develop a marketing strategy by providing advice to consumers.
3. Implementation of the Apriori Algorithm in system data begins with inputting daily sales data into a MySQL database, the biggest amoun of data, is more difficult to create tabular tables. The tabular table is then connected to the php programming, and starts building support and confidence and then will produce a final association that satisfies support and confidence.

References

[1] Pane, Kartika D. "Determination of Temporal Association Rules Pattern Using Apriori Algorithm" *Indonesian Journal of Computing and Cybernetics Systems,* vo.14, no. 2, pp 159-168, 2020.
[2] Kusrini, EmhaTaufiqLuthfi, ALGORITH DATA MINING, Edition 2, Yogyakarta:ANDI, 2010, pp 25-40.

[3] A. Maulani, "Discovering Temporal Association Rules Using Apriori Algorithm (Case Study : Toko Batik Diyan Solo)," Tesis, Program Pasca Sarjana Ilmu Komputer, Universitas Gajah Mada, 2015.

[4] F.A. Santuni, "Penerapan Algoritma Apriori Untuk Penentuan Tingkat Pesanan," Jurnal Mantik Penusa Journal In Information Systems, Vol. 2, No. 1, 2018 [Online]. Available: http://e-jurnal.pelitanusantara.ac.id/index.php   /mantik/article/download/330/216. [Accessed: 20-Sep-2020].

[5] D.T Larose and C.D. Larose, "Discovering Knowledge In Data An Introduction to Data Mining", Ed.2, New Jersey, Canada, 2014.

This page is intentionally left blank