The Implementation of Data Mining for Association Patterns Determination Using Temporal Association Methods in Medicine Data

Lisna Zahrotun, Dewi Soyusiawaty, Rahma Sara Pattihua
Department of Informatics,
Faculty of Industrial Technology,
Universitas Ahmad Dahlan, Indonesia
lisna.zahrotun@tif.uad.ac.id, dewi.soyusiawaty@tif.uad.ac.id, rahmasarapattihua@gmail.com

Abstract— Clinic is one of the businesses that perform health services for people in the surrounding environment. The clinic also provides medicines that will be given to patients who conduct health checks. The problem that occurs in these clinics is that the medicine data recap is only using excel data, the purchase of medicine stocks that are conducted only based on medicine that out of stock. Based on an interview with one of the nurse at a clinic on Yogyakarta site, occassionally, there are a case that a surge of patient that running out medicine supplies, while on the other hand there are lots of medicine accumulation occurred because these medicines was not needed by the patient. This is because the clinic has not been able to predict the medicine that are often issued by the clinic. Therefore, this research aims to build a data mining program with the Temporal Association Rules method for determining the relationship between medicines which is accompanied by the date of release of the medicine.The method used in this research is Temporal Association Rules with the Apriori Algorithm to find association rules that meet the support and confidence limits, and in the testing process lift ratio is used.The results of this research are applications that able to provide information on patterns of medicine data associations and the date of medicine's release. The test results with 8186 amount of data and support value 50% and confident value 70% with lift values above 0, the patterns of association rules obtained is 6.

Keywords— *Data Mining, Apriori Algorithm, Temporal Association Rules, Lift Rasio, Medicine data*

I. INTRODUCTION

A. The Background Problem

Data Mining is a process of extracting large amounts of data [1]. Nowadays, data mining has been implemented in various fields, including in the field of business or trade, Customer Relationship Management (CRM) [2] and education [3]. One of the data mining techniques that can be used is the association method, which is the search for patterns of relationships between items. The association method is often used to analyze sales transaction data in order to find out the market conditions of what items are often bought together by consumers. Temporal association rule is a method for finding patterns of association between items related to time [4].

Clinic is one of the businesses that perform health services for people in the surrounding environment. The clinic also provides medicines that will be given to patients who conduct health checks. The problem that occurs in these clinics is that the medicine data recap is only using excel data, the purchase of medicine stocks that are conducted only based on medicine that out of stock. Based on an interview with one of the nurse at a clinic on Yogyakarta site, there are 183 types of medicines with maximum of 630 times medicine expenditure in a month. Another problem that has not been resolved is medicines that are issued simultaneously for a certain time cannot be known. For example, during the rainy season in October-December, the expenditure of cough medicine (OBH) and paracetamol relatively increased sharply by 80% compared to normal days. As for the supply of medicines in stock globally. This condition shows some medicine items that are not needed in the excessive rainy season, while for other items of medicine are needed when the rainy season appears to be less in number. This is the impact of the problems that are owned by clinics that have not been able to find information on medicines issued simultaneously at a certain time.

From the description above, in this research a pattern of associations was searched using the Temporal Association Rules method on medicine data. This research is expected to be able to find out the types of medicines that are often issued simultaneously with the time aspects in the results of association rules.

II. METHODOLOGY

A. Research Data

The data used in this research are medicine expenditure data from January to October in 2015 and 2016.

B. Data Mining

Data mining is the process of finding interesting and hidden patterns from a large collection of data stored in a database, data warehouse, or other data storage [5]. Data mining is an implicit extraction of data that was previously unknown, and is useful in providing new information. The idea is to build computer programs that are filtered automatically by the database, looking for order or pattern. A strong pattern, if found, is likely to generalize to an accurate prediction [6]. Data mining is one of the most important steps of knowledge discovery in the database process and is considered a significant subfield in knowledge management [7]. Data mining is often also

referred to as a knowledge discovery in database (KDD). Data mining stages are shown in Figure 1 [1]
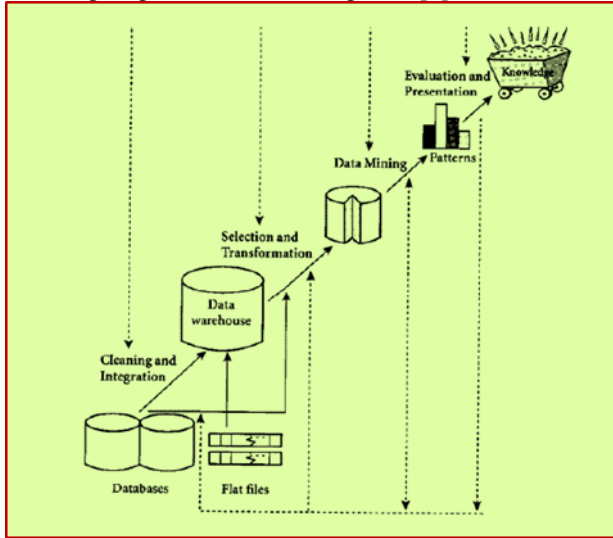

Fig 1. *Knowledge Discovery stages in Database*

C. Association Rule

The Association Rule has such applications as market basket analysis [8]. Association Rule or analysis of associations is known as one of the data mining method that becomes the basis of various other data mining methods. Especially one of the stages of association analysis that called high frequency pattern analysis (frequent pattern mining), this stage attracts many researchers to produce efficient algorithms. The importance of an associative rule can be known by two parameters, support (supporting value), namely the percentage of a combination of items in the database and confidence (value of certainty), namely the strong relationship between items in associative rules.

Association analysis is defined as a process to find all associative rules that meet the minimum requirements for support (minimum support) and minimum requirements for confidence (minimum confidence). The basic methodology of association analysis is divided into two stages:

a.   High frequency pattern analysis

This stage is looking for a combination of items that meet the minimum requirements of the value of the support in the database. The value of the 1 item support is obtained by using the equation:

$$support\ (A) = \frac{The\ amount\ of\ transactions\ contained\ A}{Total\ transaction} \quad (1)$$

b.   Formation of association rules

After all high frequency patterns have been found, then the association rules that fulfill the minimum requirement for confidence can be found by calculating the A → B associative confidence rule. The confidence value of A → B rule is obtained by the equation:

$$Confidence\ P\ (B|A) = \frac{Amount\ of\ transaction\ contained\ A\&B}{Amount\ of\ transactions\ contained\ A} \quad (2)$$

To determine the association rules to be selected, it must be sorted by Support × Confidence. Rules are taken as many as n rules that have the greatest results.

D. Apriori Algorithm

Apriori algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 to determine Frequent itemsets for Boolean association rules [4]. This algorithm tries to find all association rules that have some minimal support and some minimal trust. The biggest performance improvement in the Apriori algorithm is the reduction in the form of prospective generation sets. In the first order, all 1 itemsets are generated. For all, it will only be through itemsets which are considered as candidate itemset that were found before. The main idea is that a subset of the large itemset itself will be large. Large itemset k, is all that needed to enter the size set (k-1) [9].

Association rules for rules between items in a large transaction database have been recognized as important database research fields. The original problem dealt with in the association rule is to find a correlation between different products from a large set of data [10].

E. Temporal Association

The Temporal Association Rules can be considered as conventional association rules which may include a combination of one or more temporal relationships between items in antecedents or consequence. Based on the formulation of genuine formalism, the rules of temporal relations can be defined as follows:

Suppose I = becomes a set of binary attributes or items and T becomes a tuple database. The association rule was first proposed to be used in the transaction database, where each transaction t is recorded with an appropriate tuple [11]. The Temporal Association Rule Mining is to find valuable relationships between items in the database [12]. The Temporal Association Rules can find correlations or patterns that depend on the time between events in large volumes of data [4].

A strong Temporal Association usually occurs in the same place and at the same time while in different places it will produce a different time [12]. Data Mining can find from a series of suitable Temporal Association time transactions [13].

F. Lift Ratio

lift ratio is a measure of the strength of the association rules that have been formed. The lift ratio value is usually used as a determinant of whether the association rules are valid or invalid. To calculate the lift ratio, the equation below can be used:

$$Lift\ Ratio = \frac{Confidence\ (A,B)}{Benchmark\ Confidence\ (A,B)} \quad (3)$$

Benchmark confidence value itself can be calculated using the following formula:

$$Benchmark\ Confidence = \frac{Nc}{N} \quad (4)$$

Information:
Nc = number of transactions with consequent items
N = number of database transactions.

III. RESULT AND DISCUSSION

A. Temporal Association Stage

In this research the temporal association was carried out based on daily medicine data expenditure. So the unit of time used is the date. The calculation of support and confidence is limited to only 3 combinations of itemset.

1. High Frequency Pattern Analysis

High Frequency Pattern Analysis is used to find support values that meet the minimum support requirements. Table 1. is a set of itemset D with a time attribute that is T. There are 10 transactions involved at the time that has a lifetime. Where T = 1 is the initial time and T = 10 is the end time.

Table 1 Medicine Expenditure Data

Medicine Expenditure Transaction	Time
T1 = { Albotyl, Itamol Forte, Zenirex Syr }	1
T2 = {-}	2
T3 = { Ambroxol 30mg, Amoxcillin 500 mg, Berlosid, Betominplex, Caviplex, Cetirizine 10mg, Cairan Infus RL, Dehista, Degirol, Itamol Forte, Ikamicetin zalp, Intunal M, Masker, OBH, Paracetamol 500 mg, Supravit }	3
T4 = { Ambroxol 30mg, Berlosid, Bisolvon, Betominplex, Cetirizine 10mg, Dehista, Degirol, Itamol tab 500mg, Intunal M, Methylprednisolone, Norit, OBH, ORA, Paracetamol 500 mg, Supravit }	4
T5 = { Ambroxol 30mg, Asam Mefenamat 500gr, Betominplex, Caviplex, Cetirizine 10mg, Ciprofloxacin, Degirol, Kalium, Methylprednisolone, Norit, Neurodex tab, OBH, Paracetamol 500 mg, Simvastatin }	5
T6 = { Ambroxol 30mg, Antalgin, Amlodipine 5 mg, Bisolvon, Betominplex, Captopril 12,5 mg, Cetirizine 10mg, Degirol, Intunal M, Itamol Forte, Kalk , Ketoconazole tablet, Neurodex tab, OBH, Simvastatin, Vitazym, Vitalog C }	6
T7 = { Amoxcillin 500 mg, Antalgin, Caviplex, CEF, Degirol, Degirol, Itamol Forte, Itamol Tab, Intunal M, Methylprednisolone, Paracetamol 500 mg }	7
T8 = { Berlosid, Betominplex, Itamol Tab, Neurodex tab }	8
T9 = { - }	9
T10 = { Ambroxol 30mg, Berlosid,, BETA, Betominplex, Caviplex, Cetirizine 10mg, Degirol, Domperidone, Gentamicin 0,1%, Itamol Forte, Ikamicetin zalp, Intunal M, Methylprednisolone, OBH, Omeprazole 20 mg, Salbutamol 2mg }	10

From Table 1 there will be a support value for combinations of 1-itemset, 2-itemset and 3 itemset. The results of the 3 itemset rules from table 1 are shown in Table 2

Table 2 the result of 3 itemset rule

Combination	support (%)	Confidence
If **Betominplex** and **Cetirizine 10mg** then **OBH**	75%	83.33%
If **Ambroxol 30mg** and **Cetirizine 10mg** then **OBH**	62.5%	100%
If **Ambroxol 30mg** and **Betominplex** then **OBH**	62.5%	100%
If **Ambroxol 30mg** and **Betominplex** will be **Cetirizine 10mg**	62.5%	100%

The results of the calculation examples with apriori algorithm on the medicine sales transaction data with the highest support and confidence are Ambroxol 30mg, Betominplex, Cetirizine 10mg, and OBH.

2. Temporal Association Rules Process

After finding frequent 3-itemset that meets the minimum support of 60% and minimum confidence of 80%, namely BET-CET-OBH, AMB-CET-OBH, AMB-BET-OBH and AMB-BET-CET. Association rules that have been obtained are then processed further by adding information about the time that is in the form of transaction date, so that the resulting association rule has an interval time. The mapping results of each itemset that are valid on the transaction date and the number of occurrence frequencies are shown in Table 3.

Table 3 Results of itemset mapping that has a lifetime

No	Item	Medicine expenditure date	Occurrence of lifetime
1	AMB	[3,4,5,6,10]	[3,4,5,6,10}
2	BET	[3,4,5,6,8,10]	[3,4,5,6,10}
3	CET	[3,4,5,6,10}	[3,4,5,6,10}
3	OBH	[3,4,5,6,10}	[[3,4,5,6,10}

After finding the number of each item's occurrence frequency and the month of transaction, the next step, from the combination of rules containing itemsets in Table 3 is combined with Table 2. So that the results of the temporal association are shown in Table 4.

Table 4 Temporal Asosiasion Rule Result

BET,CET,OBH(75%.83,3%[3,4,5,6,10])
AMB,CET,OBH(83,3%.100%[3,4,5,6,10])
AMB,BET, OBH(83,3%.100%[3,4,5,6,10])
AMB, BET,CET(83,3%.100%[3,4,5,6,10])

B. Implementation

1. Import data interface

On the data import interface page, the admin is able to import medicine expenditure data and delete import results. Import data is shown in Figure 2.
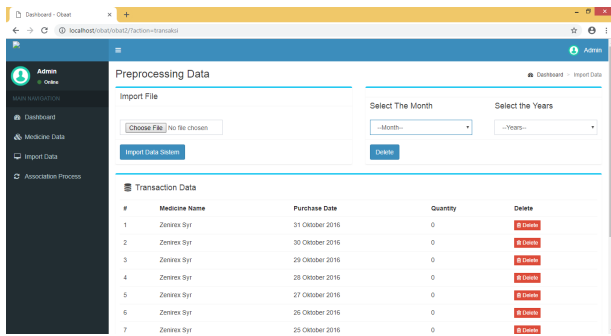
Fig. 2. Implementation of Association Page Interface

2. Process Association page

On the admin association process interface page, the admin is able to input the start date of the expenses and the end date of medicine expenditure, support and confidence. The implementation of the association process is shown in Fig. 3. As for the 2-itemset association results shown in Fig. 4, the results of the 3 itemset combination association are shown in Fig. 5 and itemset temporal association result page are shown in Fig. 6.
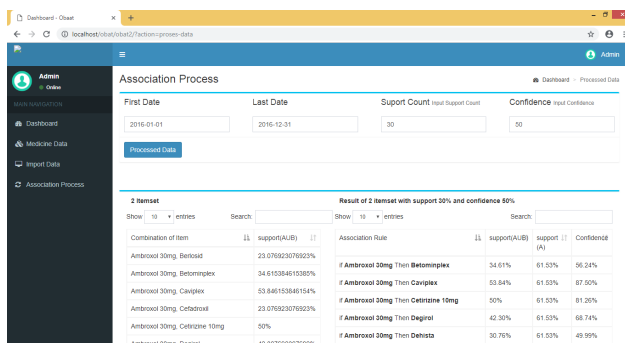

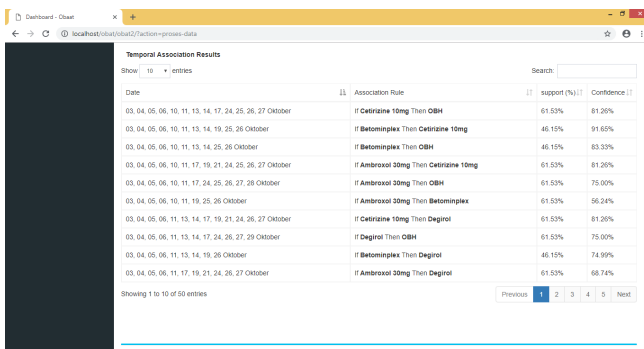Fig. 3. Association **Process** Implementation

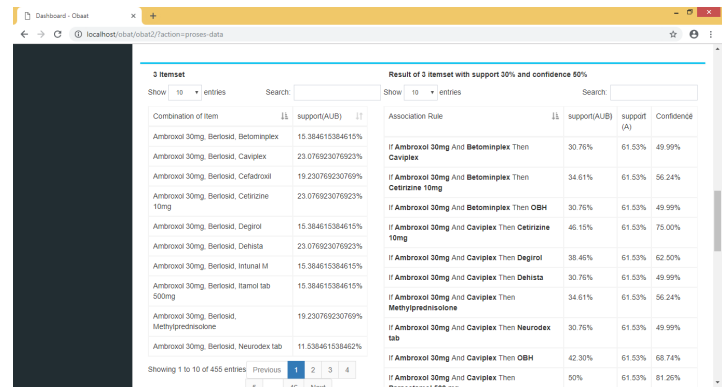Fig. 4. *2-itemset association result page*

Fig. 5. 3-*itemset association result page*

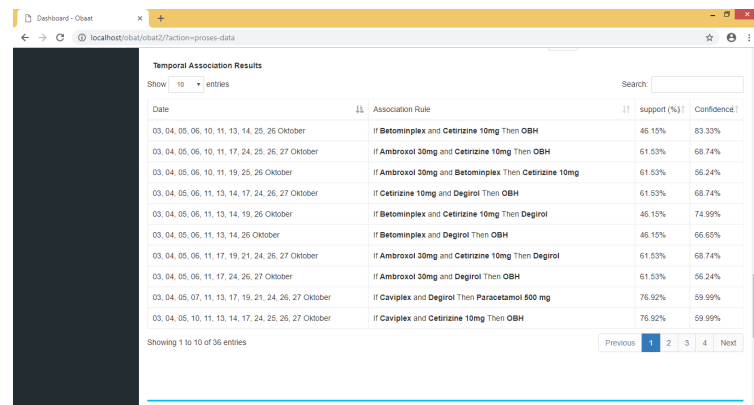And temporal association results is shown on Figure 6.

Fig. 6 *itemset temporal association result page*

C. Testing

The testing process is carried out with Transaction data 01 January 2015 - 20 October 2015 and 01 January 2016 - 20 October 2016 with a comparison between confidence for a rule divided by benchmark confidence, which can also be called support from consequence. The result can be seen in Table 5

Table 5 lift rasio testing result table

Support	Confidence	Benchmark	Lift rasio
50%	70%	71,85	1,02
50%	70%	74,06	1,05
50%	70%	77,03	1,10
50%	70%	70,30	1,00
30%	50%	50,77	1.01
30%	50%	53,89	1.07

D. Pattern Evaluation

All rules generated in data mining are then evaluated to get association rules. Evaluation is done by looking for strong rules from all existing evaluations. Rules are determined based on support and confidence values that meet the minimum limit. In this research itemset is limited to 3 itemsets considering the results of itemset search are obtained that with 3-itemset can meet support and confidence.

The value of confidence in 2016 was set at 50% and

30% support because when a maximum search of support and confidence was carried out in 2016 that met smaller support and confidence compared to 2015. Association pattern result in 2015 and 2016 are shown in Table 6 and Table 7.

Table 6. Association Pattern Result in 2015

Rules	The dates	Month
If Betominplex and Cetirizine 10mg are issued, then it will release OBH	02 , 05, 06, 07, 08, 14, 15, 16, 23, 26, 27, 29	January
	02 , 05, 09, 11, 16, 17, 27	February
	02 , 03, 04, 05, 06, 09, 13, 16, 19, 24, 26, 27, 31	March
	02, 06, 07, 08, 09, 10, 13, 15, 17, 20, 21, 24, 27	April
	03 , 04, 05, 08, 09, 10, 11, 15, 16, 17, 18, 19, 22, 24, 25, 29, 30	June
	01 , 02, 13, 15, 24, 27, 28	July
	03 , 10, 19, 21, 25, 27, 31	August
	01 , 02, 03, 04, 07, 08, 11, 14, 16, 17, 18, 21, 22, 26	September
	02 , 05, 06, 12, 13, 16, 20	October
If Betominplex and Cetirizine 10mg are issued, then it will release Degirol	02 , 05, 06, 07, 08, 12, 14, 15, 16, 23, 26, 27, 30	January
	05 , 06, 09, 16, 23	February
	02 , 03, 04, 05, 06, 09, 13, 16, 17, 18, 19, 24, 26, 27, 30, 31	March
	02, 06, 07, 08, 09, 13, 14, 17, 21, 27	April
	03 , 04, 05, 08, 09, 10, 11, 15, 16, 17, 18, 19, 22, 23, 24, 25, 29, 30	June
	01 , 02, 03, 13, 23, 24, 27, 28, 29	July
	03 , 04, 10, 18, 19, 21, 27, 31	August
	01 , 03, 04, 07, 08, 10, 11, 14, 17, 21, 22, 28, 30	September
	02 , 06, 12, 13, 15, 16, 19, 20	October
If Betominplex and Caviplex are issued then it will release Cetirizine 10mg	02 , 05, 06, 07, 12, 14, 15, 16, 23, 26, 27, 30	January
	02 , 05, 06, 09, 11, 16, 17, 23, 27	February
	02 , 03, 04, 05, 06, 09, 13, 16, 17, 18, 19, 24, 26, 27, 30, 31	March
	02, 06, 07, 08, 09, 13, 15, 17, 20, 27	April
	01 , 03, 05, 08, 09, 10, 15, 16, 19, 22, 23, 24, 29, 30	June
	02 , 03, 13, 15, 23, 24, 27, 28, 29	July
	04 , 10, 14, 18, 21, 25, 26, 27, 31	August
	01 , 02, 04, 07, 08, 10, 11, 14, 15, 16, 17, 18, 21, 22, 28, 29, 30	September
	02 , 05, 06, 12, 15, 16, 19, 20	October
If Betominplex Cetirizine 10mg are issued, then it will release Intunal / Forte	06 , 07, 08, 12, 14, 15, 16, 23, 30	January
	02 , 05, 11, 16, 27	February
	02 , 03, 04, 05, 06, 09, 13, 16, 17, 18, 19, 24, 26, 27, 30	March
	06, 09, 13, 14, 17, 21, 24, 27	April
	03 , 04, 05, 08, 09, 10, 11, 15, 16, 17, 18, 19, 22, 23, 25, 29, 30	June
	01 , 02, 03, 13, 15, 23, 24, 27, 28, 29	July
	03 , 04, 10, 14, 18, 19, 21, 24, 25, 26, 27, 28, 31	August
	01 , 02, 03, 04, 07, 08, 09, 10, 11, 14, 15, 16, 17, 18, 21, 22, 28, 30	September

Table 7. Association Pattern Result in 2016

Rules	The dates	Month
if Betominplex and Caviplex are issued, then it will be release Intunal M	04 , 05, 06, 07, 08, 09, 11, 12, 13, 14, 25, 27, 28, 29	January
	01 , 03, 05, 15, 17, 29	February
	02 , 03, 07, 08, 10, 14, 16, 18, 22, 24, 30, 31	March
	02 , 10, 11, 16, 17, 19, 24, 27, 30, 31	May
	02 , 27	June
	03, 15, 18, 20, 24, 25, 29	August
	05 , 06, 07, 08, 15, 16, 19, 26, 28, 29, 30	September
	03 , 04, 10, 11, 13, 14	October
If Betominplex and Caviplex are issued, then it will release Cetirizine 10mg	04 , 05, 06, 07, 08, 09, 11, 12, 13, 14, 16, 20, 22, 25, 26, 27	January
	01 , 05, 09, 15, 18, 22, 29	February
	02 , 07, 10, 14, 15, 16, 17, 22, 23, 24, 28, 29, 30, 31	March
	04 , 05, 06, 07, 11, 12, 13, 19, 27, 28	April
	02 , 09, 10, 16, 17, 18, 19, 23, 24, 31	May
	02 , 10	June
	25, 29	August
	02 , 05, 06, 07, 08, 15, 16, 19, 26, 28, 29, 30	September
	03 , 04, 05, 10, 11, 13, 14, 19	October

Representation of knowledge or patterns of association from the selection produced 6 patterns, namely:
1. If Betominplex and Cetirizine 10mg are issued, it will release OBH at most in June and none in May 2015

2. If Betominplex and Cetirizine 10mg are issued, it will release Degirol at most in June, and none in May 2015.
3. If Betominplex and Caviplex are issued, it will release Cetirizine 10mg at most in September and none in May 2015.
4. If Betominplex and Cetirizine 10mg are issued, it will release Intunal/Forte at most in September and none in May and Oktober 2015.
5. If Betominplex and Caviplex are issued, it will release Intunal M at most in January and none in April and June 2016.
6. If Betominplex and Caviplex are issued, it will release Cetirizine 10mg at most in January, and none in July 2016.

IV. CONCLUSION AND SUGGESTION

A. Conclusion

The conclusions that can be drawn from the results of determining the pattern of associations using the Temporal Association process are:
1. An application to determine the pattern of associations in medicine data that displays the results of the association rule value, Temporal Assication and lift ratio can be developed.
2. The test results from the Temporal Association method with the lift ratio obtained the validity of the lift value ratio above 0, that is 6 patterns with the number of transactions tested as much as 8186 and 70% confidence value with 50% support in 2015 while the 30% confidence value with support 50% in 2016.

B. Suggestions

The sales transaction data obtained does not contain the transaction hours attribute. It is expected that the next research will be able to be more detailed because it can find out the shopping pattern based on morning, afternoon or night. So that the Temporal Rules generated in addition to the transaction date also have transaction times.

REFERENCES

[1] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kauffman, 2006.

[2] L. Zahrotun, "Implementation of Data Mining Technique for Customer Relationship Management (CRM) on Online Shop Tokodiapers . com With Fuzzy C-Means Clustering," in *2017 2nd International Conferenceson Information TEchnology, Information System, and Electrical Engeenering (ICITISEE) yogyakarta, Indonesia*, 2017, pp. 298–302.

[3] H. Yuliansyah and L. Zahrotun, "Designing web-based data mining applications to analyze the association rules tracer study at university using a FOLD-growth method," *Int. J. Adnvanced Comput. Res.*, vol. 6, no. 27, pp. 215–221, 2016.

[4] E. Winarko and J. F. Roddick, *Discovering Richer Temporal Association Rules from Interval-based Data : Extended Report*, no. March. 2005, pp. 1–9.

[5] P. A. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. 2006.

[6] T. Marnoto, "Drying of Rosella (Hibiscus sabdariffa) Flower Petals using Solar Dryer with Double Glass Cover Collector," *Int. J. Sci. Eng.*, vol. 7, no. 2, pp. 155–160, 2014.

[7] T. Silwattananusarn and A. P. Kulthidatuamsuk, "Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012," *Int. J. Data Min. Knowl. Manag. Process*, vol. 2, no. 5, pp. 13–24, 2012.

[8] K. S. Kumar and R. M. Chezian, "A Survey on Association Rule Mining using Apriori Algorithm," *Int. J. Comput. Appl.*, vol. 45, no. 5, pp. 47–50, 2012.

[9] B. R. Agarwal, R. Srikant, and M. A. Ahmad, "' Fast Algorithms for Mining Association Rules ,'" *World*.

[10] H. Bathla and M. K. Kathuria, "Association Rule Mining : Algorithms Used," *Int. J. Comput. Sci. Mob. Comput.*, vol. 4, no. 6, pp. 271–277, 2015.

[11] C. P. Rainsford, *Accommodating Temporal Semantics in Data*, no. November. 1999.

[12] Z. Liang, T. Xinming, L. Lin, and J. Wenliang, "Temporal Association Rule Mining Based On T-Apriori Algorithm And Its Typcal Application," 2002.

[13] Y. Li, P. Ning, X. S. Wang, and S. Jajodia, "Discovering calendar-based temporal association rules q," vol. 44, pp. 193–218, 2003.