Edisi. .. Volume. ... Bulan 2016.. ISSN : 2089-9033

IMPLEMENTATION OF DATA MINING USING ASSOCIATION RULE METHOD ON TRANSACTION DATA IN H. APUD RESTAURANT

Rizky Muhamad Asy'ad[1], Alif Finandhita[2]
[1]Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : rizky240003@gmail.com[1] , finandhita@gmail.com[2]

ABSTRACT

H. Apud restaurant is a restaurant that became one of the pioneers in popularizing the typical cuisine of Cirebon in Cirebon area. In its service system, the restaurant offers a variety of food and beverages to the consumer and in its order, the consumer can choose the menu more than one so that the satisfaction and service of the consumer can be assured. However, there is a problem encountered business on its sale. From, the number of menus that are sold, there are several menus, which are less interesting menu for the consumers. That is why rises the idea of the owners to run a sales strategy in the form of packets. But until now, the restaurant has not had a suitable menu package because the number of data transactions that occurs make it difficult to make the package menu is right on target sale.

Data Mining is a series of processes for adding additional value of a data set. In application of data mining is used to examine each line of the transaction data to find a new pattern and can be used to determine to a decision, and then to the methods used in the implementation of data mining it is the association rule that direct to find association rules in determining the final result, and the algorithm used is FP -Growth.

Testing has been done on the test results, functionally with black box method. Based on these results it can be concluded that by implementing a data mining association rule method can help the restaurant in finding patterns of consumer purchases. It is expected that the restaurant can make a list of food and drink menu packages that can be offered to consumers that is obtained from the resulting rule.

Keywords : Association Rules, Data Mining, FP-Growth

1. INTRODUCTION

1.1 Background

H. APud restaurant is a restaurant that became one of the pioneers in popularizing the typical cuisine of Cirebon in Cirebon area. The restaurant is located at Jl. Ir. H. Djuanda Cirebon sells a variety of food and beverage menus that can be offered to consumers and obtain many transactions. But the home side has about utilizing the transaction data and just kept on file, so that the collected data cannot be put to good use to develop the business into a more advanced future.

In the home meal service system offering a variety of food and beverages for the consumer and the consumer can choose the menu, ordering more than one so that satisfaction and service to consumers can be assured. Based on interviews with the owners, there are the problems faced in running the business of selling it, from the number of menus that are sold there are several menus that are less in demand by consumers. With the onset of the matter, came the idea of the owners to run a sales strategy in the form of packets. Homeowners eat at the party who have full responsibility for making a business plan, which expects the strategy penjulan a package menu of food and beverages that can be offered to consumers, so the sale of the menu can be optimized and the level of satisfaction and service can be assured. But until now, the restaurant has not had a corresponding list of menu packages, because the number of data transactions that occur make it difficult to make the owner of the menu package on target

The availability of data that a lot and has a pattern of similar data in any sales transactions that occur can be taken into consideration in the use of the use of Data Mining. Data Mining is a series of processes for adding additional value of a set of data in the form of knowledge that had been unknown manually, using the Association Rules. Association rule mining is a technique for finding associative rules between the komnbinasi item. And algorithms used in, association rule mining to be matched so that the process can work well and quickly. To get a combination of items, the data will be dimmed using FP-Growth algorithm. Thus, it is expected with this application can facilitate the use of sales data and the determination of menu packages of food to determine the steps - steps business in Eating H. Apud.

Objectives achieved from this study include:
1. Provide new information to the manager of the restaurant to find the purchasing patterns of consumers.

2. Assist the restaurant in determining the formation of packages of food and beverages menu

1.2 Data Mining

Data mining is a series of processes for extracting value-added form of information that had been unknown manually from a database by extracting patterns from data with the aim to manipulate the data becomes more valuable information, obtained by extracting and recognizing patterns important or interesting from the data contained in the database.

Data mining is also known as Knowledge Discovery (mining) in databases (KDD), extraction of knowledge (knowledge extraction). In general, the definition of data mining can be defined as follows :

1. The process of discovery of interesting patterns from data stored in large quantities.
2. Extraction of the information useful or interesting (non-trivial, implicit, previously unknown potential usefulness) patterns or knowledge of the data that is stored in large quantities.

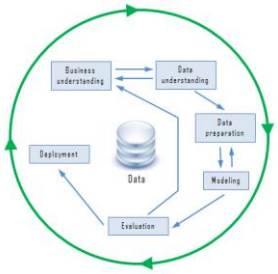
3. Exploration of automatic or semiautomatic analysis of the data in large numbers to search for meaningful patterns and rules.

1.3 CRISP-DM Methodology

CRISP-DM (Cross-Industry Standard Process for Data Mining) is standardizing data mining compiled by the three initiators of the data mining market. Namely Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR. Then developed in various workshops (between 1997-1999). More than 300 organizations that contribute in this modeling process and finally CRISP-DM 1.0 was published in 1999.

Based data mining process CRISP-DM consists of 6 phases, namely:

1. Business Understanding
2. Data Understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment



Picture 1. CRISP - DM

1.4 Association Rules

Association rule is one of the main techniques or procedures in Market Basket Analysis to look for relationships between items in a dataset and display forms of association rule. Association rule (association rule) 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 itemset first. Frequent itemset is a set of items that frequently appear simultaneously. After all patterns of frequent itemset is found, then look for a rule or rules associative linkages are eligible who have been determined.

If it is assumed that the goods sold in supermarkets is the universe, then every item will have a boolean variable that will indicate its existence or not the goods are in one transaction or a shopping cart. Boolean patterns obtained are used to analyze items purchased simultaneously. The pattern is defined in an association rule. For example, consumers typically buy coffee and milk will be shown as follows:

Coffee → Milk [support = 2%, confidence = 60%]

Association rule required a variable size specified by the user to determine the extent to which the limitation or how much output the desired user. Support and Confidence is a measure of trust and usefulness of a pattern that has been found. Support value 2% suggests that the overall total transactions consumers buy coffee and milk simultaneously as many as 2%. While the 60% confidence that showed that consumers buy certainly buy coffee and milk by 60%. Association rule has two stages of processing, namely:

1. Finding the most frequent combination of an itemset.
2. Define condition and result (for conditional association rule).

In determining a association rule, there is an interestingness measure (measure of confidence) that didaapt of the data processing with certain calculations. There are generally two sizes, namely:

1. Support: a measure that shows how much the level of dominance an item / itemset of the overall transaction. This measure determines whether an item / itemset feasible to look for his confidence (eg, of the total transactions there are, how big the dominance of an item indicates that the item A and B items purchased together).
2. Confidence: a measure that shows the relationship between the two items is conditional (eg, calculating the probability of how often the items purchased by the customer and if the customer is buying an item A).

The two measures will be useful in determining the strength of a pattern by comparing the pattern with a

minimum value of both parameters specified by the user. If a pattern of two values meet minimum pre-set parameters, then the pattern can be termed as interesting rule or strong rule. The basic methodology association analysis is divided into two stages.

1. Analysis of high frequency pattern

This stage looking for a combination of items that meet the minimum requirements of the value of the support in the database. Support an item value is obtained by the following formula

Support(A) = Sum of Transaction Containing (A) / Transaction Total … (1)

Meanwhile, the support of 2 items obtained from the following formula.

$$Support\,(A,B) = P(A \cap B) \quad \ldots (2)$$

2. Formation of association rules

After all the high frequency pattern is found, then search for the associative rules that meet the minimum requirements for confidence by calculating confidence associative rule A -> B of the support of high-frequency pattern A and B, using the formula:

$$Confidence = P(B \mid A) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{Jumlah\ transaksi\ yang\ mengandung\ A}$$

1.5 FP-Growth Algorithm

FP-Growth is a frequent itemset search algorithm derived from the FP-Tree by exploring Tree from the bottom toward the top. FP-Growth algorithm is an alternative algorithm that is quite effective to locate the data set that appears most frequently (frequent itemset) in a large data set. FP-Growth algorithm is an algorithm Association Rules are quite often used. FP-Growth algorithm was developed from a priori algorithm. Priori algorithm produces a combination that is very much so very inefficient. FP-Growth algorithm is one solution of a priori algorithm that takes a very long time because it had to do pattern matching that repeatedly. While in the process of FP-Growth Algorithm there are many advantages that proved to be very efficient because only the mapping of data or database scan 2 times to build up the structure of "Tree". This algorithm determines the frequent itemset ending in a particular suffix using the divide and conquer method to break down the problem into smaller sub problems.

FP-Growth methods can be divided into three main stages, namely as:

1. Phase generation of conditional pattern base
2. Phase generation of conditional FP-Tree
3. Phase frequent itemset search

2. CONTENT RESEARCH

1.1 Data extraction

Data extraction is performed in this stage is making the sales transaction data is carried out from data derived from the database. Extracted data is stored in a database and then followed by the data processing to the next stage. Here's an example of a sample of transactions with 46 transactions contained in Eating H. Apud where in the data is another among Id, Ticketid, Menuitemid, MenuItemNam.

1.2 Selection Attributes

Based on the information you want to get by users regarding food and drink anything purchased simultaneously by consumers, so at this stage the attributes that will be used is Ticketid, MenuItemName which aims to select what data will be used during the process of mining and the necessary database for purposes of analysis, since not all of the attributes used during the mining process.

1.3 Cleaning Data

Data cleaning eliminates noise and eliminate irrelevant data or inconsistent. Omitted data is data on a transaction that there is only one menu purchase, because the data does not have a relationship of association with another menu. Because no amount of cleaning time of the transaction there is only one menu purchases in one transaction then do not pemberishan.

2.4 Selecting Modeling Techniques

By using Data Mining, transaction data will be processed by the rules of the association to find patterns of buying menus often purchased by consumers. In this stage of the method used is Assosiacion Rule.

This study aims to find patterns of consumer purchasing associations. The modeling techniques used in this research is the method of Association Rules while the algorithm used in this study is the algorithm FP-Growth By applying data mining, transaction data will be processed by the association rules or methods Association Rules for finding patterns of purchases are often purchased by consumers.

Table 1. Results of Investigation and the Elimination MenuItemid

Ticketid	*Menuitemid*
3.445	4,16,13
3.446	14,4,16,3,9,6,13,18
3.447	14,4,3,9,7
3.448	14,4,16,3,6,7

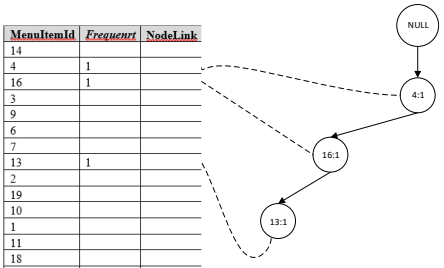
3.449	14,4,16,3,9,10
3.450	14,4,16,19,10
3.451	14,3,13,10,1
3.452	14,4,3,9,2
3.453	14,4,16,3,6,2
3.454	14,4,9
3.455	14,16,3,9,7
3.456	6,7,13
3.457	4,16,3,9
3.458	14,4,3,19
3.459	14,4,16,3,6,7,13,1
3.560	14,4,16,6,7,2
3.561	14,4,16,3,9,7
3.562	14,16,9,17
3.563	14,4,9
3.564	6,2,5,17
3.565	4,16,7
3.566	14,4,16,9,18
3.567	14,13,2
3.568	14,16,3,9,6,7
3.569	14,3
3.570	14,4,16,3,7,5,12
3.571	14,4,9,6
3.572	16,2
3.573	14,4,16,19
3.452	14,4,3,9,2
3.453	14,4,16,3,6,2
3.454	14,4,9
3.455	14,16,3,9,7
3.456	6,7,13
3.457	4,16,3,9
3.458	14,4,3,19
3.459	14,4,16,3,6,7,13,1
3.560	14,4,16,6,7,2
3.561	14,4,16,3,9,7
3.562	14,16,9,17
3.563	14,4,9
3.564	6,2,5,17
3.565	4,16,7
3.566	14,4,16,9,18
3.567	14,13,2
3.568	14,16,3,9,6,7
3.569	14,3
3.570	14,4,16,3,7,5,12
3.571	14,4,9,6
3.572	16,2
3.573	14,4,16,19
3.452	14,4,3,9,2
3.453	14,4,16,3,6,2
3.454	14,4,9
3.455	14,16,3,9,7
3.456	6,7,13
3.574	4,16,1
3.575	13,2
3.576	14,4,16,3,9,6,7
3.577	4,16,3
3.578	14,4,16,6
3.579	14,4,16,9
3.580	14,4,9
3.581	4,16,3,9,13,19
3.582	14,16,3,6,12
3.583	14,4,16,3,10,18
3.584	14,16,3,9,6,19,12
3.585	14,3,1
3.586	14,4,2
3.587	14,4,10,5
3.588	13,17
3.589	14,4,16,3,9,6,7,13,1
3.590	14,4,16,6,9,19

2.5 *FP-Growth* Analyst

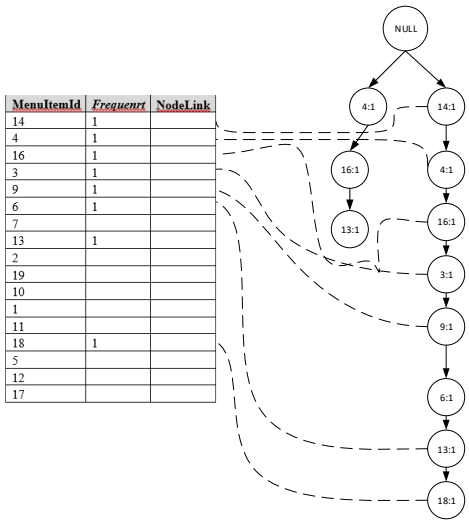
After getting the data that has been sorted, the next thing to make tree. Here are the stages in the manufacture of tree:

1. Create a Tree with Root who had no children.
2. If Root has no children, then the data that appears first used as the first child of the Root accompanied by support count = 1.
3. If the Root already have children, but the data which appears different from the children of Root, then the data is the next child of the Root and support its count incremented by 1.
4. If the Root already have children, and the data that appears together with the support of the child increases its count 1.
5. If the existing data that appears in the notation son, then his support count incremented by 1.

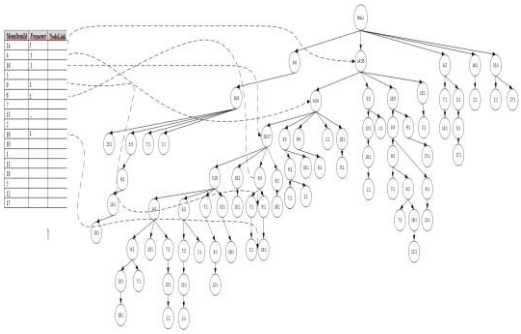
MenuItemId	*Frequent*	NodeLink
14		
4	1	
16	1	
3		
9		
6		
7		
13	1	
2		
19		
10		
1		
11		
18		
5		
12		
17		



Picture 2 *FP-Tree* from *Ticketid* 3.445



Picture 3 *FP-Tree* from *Ticketid* 3.446



Picture 4 *FP-Tree* from *Ticketid* 3.590

After the stage of development of the FP-Tree from a set of transaction data, will be applied to FP-Growth algorithm to search Frequent itemset significant. FP-Growth algorithm is divided into three main steps, namely :

1. Generated Conditional Pattern Base
   Conditional Pattern Base is subdatabase containing the prefix path (path prefix) and the suffix pattern (pattern suffix). Generation conditional pattern base obtained through the FP-Tree that had been established earlier.

Table 2. *Conditional Pattern Base*

Suffix	Conditional Pattern Base
17	{{14,16,9:1},{6,2,5:1},{13:1}}
12	{{14,4,16,3,7,5:1},{14,16,3,9,6,19:1},{14,16,3,6:1}}
5	{{14,4,16,3,7:1},{14,4,10:1},{6,2:1}}
18	{{14,4,16,3,9,6,13:1},{14,4,16,3,10:1},{14,4,16,9:1}}
1	{{4,16:1},{14,4,16,3,6,7,13:2},{14,3,13,10:1}}
19	{{4,16,3,9,13:1},{14,4,16:2},{14,4,3:1},{14,16,3,9,6:1},{14,16, 9:1}}
10	{{14,4,16,3,9:1},{14,4,16,19:1},{14,3,13:1}}
13	{{4,16,3,9:2},{14,4,16,3,9,6:1},{14,4,16,3,9,7:1}
	{14,4,16,3,6,7:1},{14,3:1},{6,7:1}}
2	{{14,4,16,3,6:1},{14,4,16,6,7:1},{14,4,3,9:1}, {6:1},{16:1},{13:1}}
7	{{4,16:1},{14,4,16,3,9,6:1},{14,4,16,3,6:2} ,{14,4,16,6:1},{14,4,3,9:1},{14,16,3,9,6:1},{6:1}}
6	{{14,4,16,3,9:2,{14,4,9:1},{14,16,3,9:2}}
9	{{4,16,3:2},{14,4,16,3:5},{14,4,3:2},{14,16,3:3}}
3	{{4:16:3},{14,4,16:10},{14,16:4}}
16	{{14,4:16}}
4	{14:24}
14	{{14:35}}

Table 3 . *Conditional Frequent Pattern Tree*

Item	Conditional FP Tree
12	{12,14:3}{12,16:3}{12,3:3}
18	{18,4:3},{18,16:3},{18,14:3}
1	{1,14:3},{1,4:3},{1,16:3},{1,3:3}, {1,13:3}
19	{19,14:5},{19,16:5},{19,4:4} ,{19,3:3},{19,9:3}
10	{10,14:5},{10,4:4},{10,16:3},{10,3:3}
13	{13,4:5},{13,16:5},{13,3:5},{13,14:4} ,{13,9:3},{13,6:3},{13,7:3}
2	{2,14:4},{2,4:4},{2,16:3}
7	{7,14:8},{7,4:8},{7,16:8},{7,3:7},{7,9:5} ,{7,6:6}
6	{6,14:11},{6,4:8},{6,16:10},{6,3:8},{6,9:5}
9	{9,14:14},{9,4:11},{9,16:9},{9,3:12}
3	{3,14:20},{3,4:16},{3,16:19}
16	{16,14:23},{16,4:16}
4	{4,14:24}

Table 4. Result *Frequent Pattern Generated*

Item	Conditional FP Tree
12	{14,16,3:3}
18	{4,16,14:3}
1	{14,4,16,3,13:3}
19	{14,16:5},{4:4},{3,9:3}
10	{14:5},{4:4},{16,3:3}
13	{4,16,3:5},{14:4},{9,6,7:3}
2	{14,4:4},{16:3}
7	{14,4,16:8},{3:7},{9:5},{6:6}
6	{14:11},{4:8},{16:10},{3:8},{9:5}
9	{14:14},{4:11},{16:9},{3:12}
3	{14:20},{4:16},{16:19}
16	{14:23},{4:16}
4	{14:24}

Because the value of the specified confidence of 70%, then that does not meet these limits will be eliminated. Then the rules can be obtained as follows:
:

Table 5. Result *Generate Confident*

Rule	***Minimum Confidence***
9→3, 14	3/3*100=100%
9 → 4, 14	3/3*100=100%
6 → 3, 14	3/3*100=100%
6 →16, 14	3/3*100=100%
6 →16, 3	8/8*100=100%
6 →16, 3, 14	4/5*100=80%
6 →4, 14	4/5*100=80%
6 →4, 16	9/12*100=75%
6 →3, 16, 14	10/12*100=83,3%
7 → 3, 14	9/12*100=75%
7 →16, 14	14/19*100=73,6%
7 →16, 3, 14	3/3*100=100%
7 →3, 14	3/3*100=100%
7 →3, 16, 14	3/3*100=100%
1 →3, 14	3/3*100=100%
10 → 3, 14	8/8*100=100%
19 →3, 14	4/5*100=80%
12 →16, 14	4/5*100=80%
12 →3, 14	9/12*100=75%
12 →3, 16, 14	10/12*100=83,3%
18→3, 14	9/12*100=75%
18→16, 14	14/19*100=73,6%
7 → 3, 14	9/12*100=75%
7 →3, 16, 14	3/3*100=100%
1 →3, 14	3/3*100=100%
10 → 3, 14	8/8*100=100%
19 →3, 14	4/5*100=80%
12 →16, 14	4/5*100=80%
12 →3, 14	9/12*100=75%
12 →3, 16, 14	10/12*100=83,3%
18→3, 14	9/12*100=75%
18→16, 14	14/19*100=73,6%

Table 5. Result *Rule* being Package

Rule
Es Teh Manis →Empal Asem, Nasi
Es Teh Manis → Empal Asem, Nasi
Es Jeruk → E mpal Asem, Nasi
Es Jeruk → Sate Kambing, Nasi
Es Kelapa→ Empal Asem, Nasi
Es Kelapa → Sate Kambing, Nasi
Aqua →Empal Asem, Nasi
Jeruk Hangat → Sate Kambing, Nasi
Jeruk Hangat →Empal Asem, Nasi
Teh Botol→Empal Asem, Nasi
Teh Botol→ Sate Kambing, Nasi

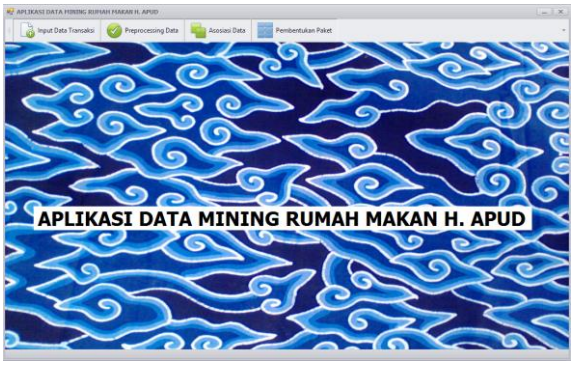
From the results of the rule that has been obtained, then the formation of packages that can be offered in the Eating H. Apud with the following provisions:
1. The number of bundled menu at least 3 to 5 drinks menu and there are one and rice.
2. Package price of at least Rp. 15,000 to Rp. 45,000
3. Package is used every 1 month

Tabel 6. Pemaketan Menu yang memenuhi Batasan Paket

Paket	Nama Menu	Harga
Paket 1	Es Teh Manis	Rp. 4.000
	Empal Asem	Rp. 25.000
	Nasi	Rp, 3.000
	Total	Rp. 32. 000
Paket 2	Es Teh Manis	Rp. 4.000
	Empal Asem	Rp. 15.000
	Nasi	Rp, 3.000
	Total	Rp. 22.000
Paket 3	Es Jeruk	Rp. 7.000
	Empal Asem	Rp. 15.000
	Nasi	Rp, 3.000
	Total	Rp. 25.000
Paket 4	Es Jeruk	Rp. 7.000
	Sate Kambing	Rp. 25.000
	Nasi	Rp, 3.000
	Total	Rp. 35.000
Paket 5	Es Kelapa	Rp. 7.000
	Empal Asem	Rp. 15.000
	Nasi	Rp, 3.000
	Total	Rp. 25.000
Paket 6	Es Kelapa	Rp. 7.000
	Sate Kambing	Rp. 25.000
	Nasi	Rp, 3.000
	Total	Rp. 35.000
Paket 7	Aqua	Rp. 4.000
	Empal Gentong	Rp. 15.000
	Nasi	Rp, 3.000
	Total	Rp. 22.000
Paket 8	Jeruk Hangat	Rp. 6.000
	Empal Asem	Rp. 15.000
	Nasi	Rp, 3.000
	Total	Rp. 23.000
Paket 9	Jeruk Hangat	Rp. 6.000
	Sate Kambing	Rp. 25.000
	Nasi	Rp, 3.000
	Total	Rp. 34.000
Paket 10	Teh Botol	Rp. 5.000
	Empal Asem	Rp. 15.000
	Nasi	Rp, 3.000
	Total	Rp. 23.000
Paket 11	Teh Botol	Rp. 5.000
	Sate Kambing	Rp. 25.000
	Nasi	Rp, 3.000
	Total	Rp. 33.000

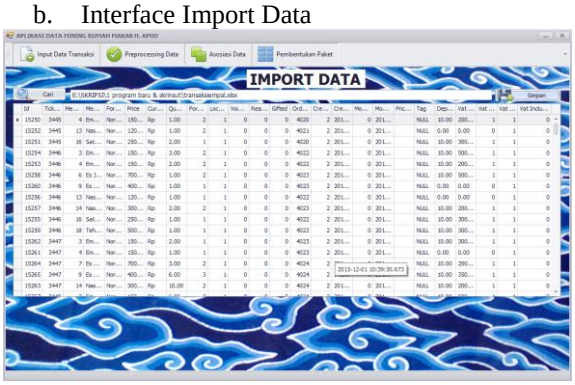
2.6 nterface Implementation

a. Interface Main Menu



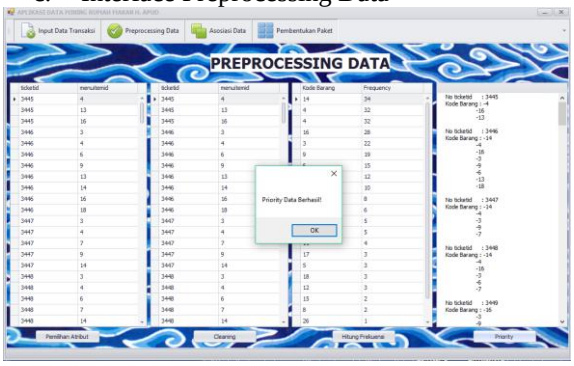
Picture 5 Main Menu Application

b. Interface Import Data



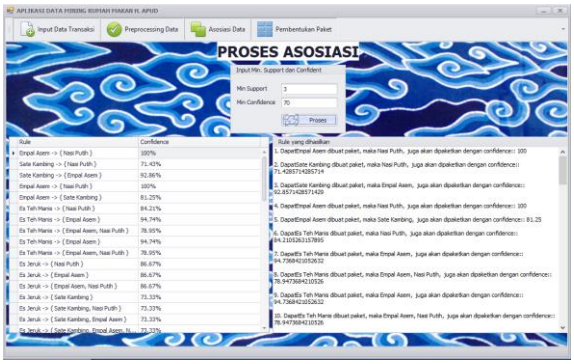
Picture 6 Interface Import Data

c. Interface Preprocessing Data



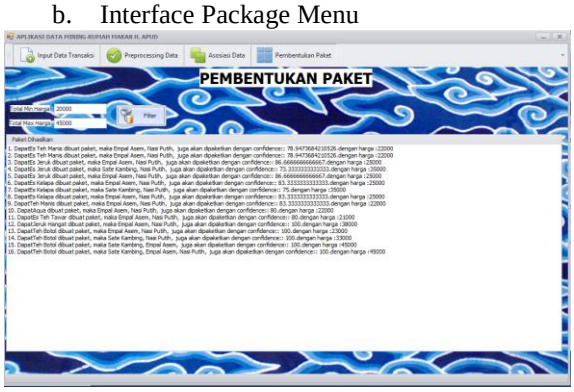
Picture 7 Interface Preprocessing Data

d. Interface Process Association



a. Picture 8 Interface Process Association

b. Interface Package Menu



Picture 10 Interface Packages Menu

3. CLOSING

Based on the analysis and testing has been done it can be concluded as follows:
1. Applications are built to provide information on the menu by consumers' purchasing patterns.
2. Applications built to help provide information about food and drink menu packages are frequently bought by consumers so that they can be considered for the formation of menu packages to consumers.

The advice given for thesis research in the development of this system are as follows:
1. Before the development of the implementation of the Data Mining built, especially at the time of import the data so that the process of extracting data can be more quickly and accurately.

Required comparison with other algorithms such as a priori or CT-Pro in the development of data mining applications that can be used as a comparison in the accuracy and efficiency of the rule results in the mining process, so that with the comparator can know which one is better algorithms and optimal development..

REFERENCES

[1] Nofriansyah, Dicky, Konsep Data Mining Vs Sistem Pendukung Keputusan, Deepublish, Yogyakarta, 2014.
[2] P. Chapman, J. Clinton, R. Keber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, CRISP-DM 1.0, Step-by-step data *mining* guide, 2000
[3] H.A.Fajar, Data Mining. Andi, 2013.
[4] Kusrini, Strategi Perancangan dan Pengolahan Basis Data, Andi, 2007.
[5] B. Santoso, Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu, 2007.
[6] Yuhefizard, Database Management Menggunakan Microsoft Access 2003, Jakarta : PT. Elex Media Komputindo, 2008.
[7] Prasetyo Eko, Data Mining, Andi, 2012.
[8] J. Han and M. Kamber, Data Mining : Concept and Techniques, 2nd ed. San Fransisco: Morgan Kauffman, 2006.
[9] Kusrini and E. T. Lutfi, Algoritma Data Mining, Andi, 2009.
[10] J. Santoni, "Implementasi Data *Mining* Dengan Metode Market Basket Analysis," *Teknologi Informasi dan Pendidikan*, vol. 5, p.2, Sep.2012
[11] D.T. Larose, DISCOVERING KNOWLEDGE IN DATA : An Introduction to DATA MINING, Hoboken, New Jersey: John Wiley & Sons, Inc., 2005.
[12] K. and E. T. Luthfi, Algoritma Data Mining, Andi, 2009