LAPORAN PRAKTIKUM
ANALISIS KEPUTUSAN DAN DATA MINING
ASSOCIATION RULE – MARKET BASKET ANALYSIS (AR-MBA)

KELOMPOK IP-17
1. Emmo Al Rasyid
2. Gusti Adli Anshari

ASISTEN DM-38

Kriteria Penilaian		Max
Format		10
Isi		50
Analisa		40
TOTAL		100

Yogyakarta,.....2015

Asisten

(.....)

LABORATORIUM DATA MINING
JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
2015

CHAPTER I
INTRODUCTION

1.1 Description of the problem

Now, Yogyakarta is not like in the past that still as traditional city, still any traditional bike everywhere, Yogyakarta now become metropolitan city, only car and motorcycle which use highway, there is so many shopping place, etc. in this city. For example of shopping place is minimarket, because so many minimarkets in Yogyakarta now, minimarket is such as supermarket but this shop is work in small scale. Minimarket always visited by customer because this shop sells some item for daily customer needs.

AR-MBA (Association Rule – Market Basket Analysis) is data mining techniques to find the associative rules between combinations of items, work to determine which attributes to be obtained simultaneously, The task of the Asociation rule is to look for rules that do not cover to measure the relationship between two or more attributes. In this study, AR-MBA research done to identify and analyze two or more items in one receipt with combination from different department of a retail business in minimarket, and the objective of this study is the researcher want to analyze about discount in minimarket in Yogyakarta, the researcher want to analyze minimarket based on 100 receipts that collected from minimarket customer with buy some product in that minimarket. The research object of the researcher is 7 Cemara, this minimarket is located at Kaliurang Street Km 5.8, Kentungan, Sleman, Yogayakarta. This minimarket established on January 2014.

Generally, the trading system in this minimarket is same like other minimarket in Yogyakarta, namely using cashier machine, when the customer will buy a product in 7 Cemara, the cashier operator will scan the barcode on each product with scanner, then the price of a product will appear on screen, after the customer pay that product, the cashier operator will push button that will print out the receipts. From first open until now, 7 Cemara always crowded by customer, the customer always buys some product in 7 Cemara, In the other hand 7 Cemara is available with @wifi.id wifi connection, so this is one of reason why 7 Cemara always crowded by customer.

The researcher wants to analyze 7 Cemara because 7 Cemara is one of new minimarket in Yogyakarta with @wifi.id wifi connection, so many customer come from morning till midnight, as a matter of fact there are some customer which stay in 7

Cemara from midnight till morning, then the researcher think that minimarket is interested to analyze about discount. As the researcher know, that minimarket not give discount with the product yet. So the researcher wants to analyze and give recommendation about discount in 7 Cemara, because if that minimarket start to give discount, it can be through increasing the number of customer per day that come to 7 Cemara.

The research start with collected 100 receipts that the researcher got from customer 7 Cemara. After got 100 receipt, then the researcher filter that receipt based on receipts requirement that needed to analyze. The requirement is the receipts must be contained 2 items or more with different department, ex: customer buys drink and snack. After the researcher collect transaction receipt and do data pre-processing. Then the researcher input the data from receipts into Microsoft Excel to divide each item into some department. Then start do calculation with Rapid Miner software, input data to this software is from data in Microsoft Excel, then add some material in that software to help calculate the data, after that add subjective support value and confident value to fulfill lift ratio requirement (lift ratio > 1). Finally the researcher can get the result, the result is show department calculation that fulfill lift ratio requirement. From it, then the researcher try to make solution from the problem with give recommendation about which product or which department that need a discount.

1.2 Problem Formulation

There are some problems formulated in this research, which are as follows:

1. How is the associative relations that happened between items in 7 Cemara minimarket?
2. How is the problems solutions for discount in 7 Cemara minimarket based on analysis of AR-MBA?

1.3 Goal of Practicum

1. Students able to know one of method association in data mining.
2. Give understanding about prosedure of Market Basket Analysis.
3. Students can process a large enough data so that the data can be used by using Association Rule.

CHAPTER II
LITERATURE REVIEW

2.1 Deductive Study

2.1.1 AR-MBA

Association in data mining is the work to determine which attributes will be obtained simultaneously. In the business world the term is commonly known as affinity analysis. The task of the association rule is to find a rule that does not cover to measure the relationship between two or more attributes.

Association rule is a form if the "previous incident" and "consequences". (IF antecedent, THEN consequent). Along with the calculation of support and confidence of rules. The pattern of association to be one of the most interesting functionality in extracting data (Kumar and Wahidabanu, 2007).

Association Rule is a data mining technique to find the association rules between combinations of items. For example, with association rules mining in purchasing analysis we can know how likely a customer buys coffee and sugar at the same time. With this knowledge owner can adjust the placement of the goods or designing a marketing campaign using a combination of discount coupons for certain items (Wiwin, 2008).

One example of application of Association Rule is Market Basket Analysis. Association Rule became known for its application to analyze the contents of the purchase shopping cart, so the Association Rule is also commonly referred to as Market Basket Analysis. Association Rule also known as one of data mining techniques that became the basis of a variety of other data mining techniques.

Market Basket Analysis is one example of the application of Association Rule. To convey the basic idea of the Market Basket Analysis, starting with a shopping cart to see the image in Figure 2.1 which contains a variety of goods purchased by someone in a supermarket. This basket contains a variety of items such as bread, milk, cereal, eggs, sugar, and so forth. A basket tell us about what is purchased by a consumer at a time. A complete shopping list that was obtained from all the information that consumers give us very much, and this may explain

what items are most important from a business sale is "what goods purchased by consumers and when".

Each consumer buys a set of different items, in different amounts, and in a different time. Market Basket Analysis uses the information of items that is purchased by consumers to provide a signal or information that is who they are and why they made the purchase? Market Basket Analysis provides an understanding of the merchandise by telling us which products are usually purchased together and which product is approved to be promoted. This information can be used for several purposes which are as follows:

1. More profitable advertising and promotion. Market Basket Analysis on advertising and promotion in order to have a better understanding regarding to how buyers respond to and communicate over the products offered, for the purpose of the retailer "How do I change this sale? What else is sold and what was advertised ".
2. More precise targeting in return ROI (Return on Investment). Market Basket Analysis is used to optimize campaigns and promotions to increase sales and margins by targeting more precise.
3. Loyalty card promotions with longitudinal analysis. Longitudinal Market Basket Analysis enables users to buy the characters retailer customer behavior over time. Retailers use loyalty cards to capture the lifecycle of data so that they can analyze the purchasing behavior of customers when shopping. For example a toy retailer explained that he did not make sense to sell a game engine (with a slight margin) except for customers who also buy the game software and accessories (high margin). They use the Market Basket Analysis of loyalty card data to determine their overall margin on sales of video games and promotions to make the memory of the customer and affect a buyer to purchase games and accessories from them and not from other retailers.
4. Determine the layout of the new store (new store layouts) or attract more traffic to the store, set of products which will be placed in a special place. Market Basket Analysis also uses the space to improve traffic planning and visual merchandising to boost sales.
5. Identify when the problem in pairs / coupon (issue coupons). To increase sales or spending items into inventory.

There are a lot of definitions of Market Basket Analysis, such as Market Basket Analysis a mathematical technique used by marketing professionals to express the similarity between the individual products or product groups. Market Basket Analysis with respect to a set of problems related to flying businesses to find out from the point of sale transaction data. Market Basket Analysis is a general term for the methodology of the study of the composition of the basket of groceries purchased by households during a time of shopping. Market Basket Analysis is a collection of a combination of products purchased together. Market Basket Analysis trend analysis of an item bought by the same customer at the same time. Market Basket Data is data that describes the transaction underlying the three different entities, namely Customers, Orders / purchases, and Items (goods)

Introduction of the customer at any time make it possible to instantly recognizable, such as the frequency of purchases made by the customer. Three levels of market basket data is important to understand the request quickly. These measurements give an idea for a business. In some cases, there are some repetitive buyers, so that the proportion of the purchase of every customer is close to 1. This suggestion is used by a company to increase sales per customers. Or the amount of each purchase of products close to 1, suggestion can be an opportunity for cross-selling during the purchase process. Important whether or not an associative rule can be determined by two parameters, support (the support) is the percentage of the combination item page and confidence in the database (the certainty) that the strength of the relationship between items in the associative rule.

So, our research held in 7 Cemara at Jalan Kaliurang km. 5,8, Kentungan, Sleman, Yogyakarta. There are 100 pieces of receipt of goods purchases that already collected. Conducting Pre Data Processing, Data Processing, Establishment of Association Rule. The application used to make this report are *Microsoft Excel, Microsoft Visio*, and *Rapid Miner* software.

2.1.2. Product Discount

According to Hoek and Roelants (1991), price discounting is a well-known merchandising technique whereby manufacturers, retailers, or both, offer consumers an economic incentive to induce them to purchase a particular brand. Most of price discounting researches has concentrated this issue with three main impacts. The impacts are as follows:

1. Impact on Market Share

   Massy and Frank (1965) investigated the short term effects of temporary price discounts and found that both brand-loyal and non-loyal buyers responded to a discount promotion. Hinkle (1965) argued that a brand's age may influence the extent to which a price discount can increase its share. He found that price discounts were most effective with new brands, which tended to achieve higher gains with smaller price reductions than more established brands. More than a decade later, Dodson, Tybout and Sternthal (1978) corroborated Hinkle's findings and concluded that price discounting increased the market share of the promoted product, at least in the short term. Furthermore, they suggested that a high discount led to a greater increase in market share than a low discount.

2. Impact on Brand Switching

   Other studies explored the subsequent behavior of brand switchers to determine whether consumers reverted to the purchase patterns they held prior to the promotion. Lawrence (1969) and Shoemaker and Shoaf (1977) concluded that this was the case and suggested the market share gained from the promotion could be as temporary as the promotion itself. Thus they concluded that these promotions may have a limited effect because they serve only to disrupt consumers' short-term purchase behaviour, which eventually resumes its normal pattern.

3. Impact on Purchase Timing and Quality

   Temporary price discounts may affect other aspects of consumers' purchase behavior, such as the quantity of product they purchase, and their inter purchase intervals. Wilson, Newman and Hostak (1979) found a strong relationship between the buying situation and the number of units purchased. Shoemaker (1979) concluded that price discounts have more effect on the

quantity purchased than on buyers' inter-purchase interval, although later research questioned this. Blattberg Eppen and Lieberman (1981), Neslin, Henderson and Quelch (1985) and Gupta (1988) concluded that these promotions may only displace sales that would have otherwise occurred at the product's usual price, thus delaying their subsequent purchase of it and competing brands. These conclusions raise an important question about the cost-effectiveness of price discounts.

4. Impact on Sales

   In accordance with Gupta's (1988) claim that the promotion after discounts is caused by purchase time acceleration and stockpiling which supports Dodson et al.'s (1978) finding about increased products market share, Hoek and Roelants (1991) mentioned that when a discount period ends, sales decline, however most sales remains above the weekly sales level recorded before promotion.

5. Impact on Competitors

   In general, sales of competitors' products declined during the discount week (Hoek & Roelants, 1991). According to Hoek and Roelants (1991), price discounts potentially beneficial effect on competitors because price promotions may not always result in increased profit for a particular brand.

In summary, manufacturers who promote their brands by way of temporary price discounts may, in the short term, induce buyers of competing brands to purchase their product, but it appears that price discounts do not usually have a permanent effect on consumers' brand preferences. Discount promotions may not necessarily have an adverse effect on sales of competing brands.

2.2 Inductive Study

Hoanca and Mock (2011) did a research about Market Basket Analysis in a university bookstore located in USA named UAA Bookstore in order to find pairs of items which provides an actual magnitude of the estimation of potential revenue increases. According to Hoanca and Mock (2011), pairs of items have two different affinities which are positive affinity and negative affinity. Pairs of items with positive affinities should be packaged together, placed in a close proximity and/or priced as a bundle while negative affinities pairs can be object of market

segmentation or increase revenues by discounting one of the items, if done properly (Vindevogel et al., 2005).

In UAA bookstore, Hoanca and Mock (2011) applied Market Basket Analysis on transactions from February up to March 2010 with a total of 13,916 transactions involving 28,462 items purchased in pairs and 4,332 items purchased individually. The items sold in UAA bookstore includes textbooks, apparel, snacks, and office products and at least 17,231 different pairs of items were sold in one transaction. However, only 88 pairs were sold together in at least 10 transactions, 250 were sold together in at least 5 transactions and 586 were sold together in at least 3 transactions. This data set was taken at the end of the academic year which gives strong affinities regarding to graduation equipment.

The data is processed with minimal support and confidence level respectively 0.03 and 0.8 which ends most pair items bought as follows:

1. Tassels and caps
2. Tassels and gowns
3. Caps and gowns
4. Master hood and tassels

This research is followed with a sensitivity analysis. As conclusion of this research the revenue obtained by UAA bookstore depends on the price sensitivity and saturation level of affinities higher sensitivity level with low saturation level gives the highest profit prediction for UAA bookstore.

On the other hand, Annie and Kumar (2012), focused on researching using the Market Basket Analysis with K-Apriori data mining algorithm in order to determine placement of goods, designing sales promotion for different segments of customers to improve customer satisfaction and the profit of a supermarket. The object of research used by Annie and Kumar (2012) is Anantha Stores which is a top supermarket in Tirumelveli city. Ananth stores are organized in eight separate sections which includes household items, fruits and vegetables, bakery, kitchen wares, toys, gifts, textiles, pharmacy. The household section is the main focus of their research and it includes more than 700 items with different brand and varying prices.

K-Apriori is an efficient association rule discovery method which uses Apriori property (Annie & Kumar, 2012). Apriori property insists that all non-empty

subsets of a frequent item set must also be frequent and the K-Apriori algorithm extracts a set of frequent item sets from the data. Then, pulls out the rules with the highest information content for different groups of customers by dividing the customers in different clusters.

As result, items I26, I34, I303, and I313 are frequently bought together in the Anantha Supermarket household item section. From the observed result, in sparse dataset like market databases, K-Apriori algorithm effectively generates highly informative frequent itemsets and association rules for Anantha Stores.

Different with the previous researches, Kaur and Singh (2013) focused on maximizing the minimum confidence level in FP-Growth (Frequent Pattern Growth) algorithm based on primary data of retail transactions in sports store. The result of this research is later on used to improve the sport store's product recommendations, promotions, product placement (layout), and marketing strategy.

The result of their research is that there are several products that must be kept on the shelves together in order to increase profits from the products which are frequently purchased together such as tumbler, towel, running shoes, socks, and backpacks.

All in all, from the previous researches, it can be inferred that association rule consist of many different algorithm but they are used for the same purpose which is association rule mining from the transactional data. This research may be used as a comparison or reference for future researcher to use because there are not many researches regarding to price discounting available on the internet.

CHAPTER III
RESEARCH METHOD

3.1 Object of Research

7 Cemara is minimarket with @wifi.id Wi-Fi connection, so many people always come to 7 Cemara to get @wifi.id Wi-Fi connection and buy some product in that minimarket. The researcher using initial condition layout and 100 receipts to identify and analyze, then try to make discount recommendation to 7 Cemara. There are profile of 7 Cemara:

Name : 7 Cemara

Address : Kaliurang Street Km 5.8 Kentungan, Sleman, Yogyakarta

Established : January 2014

Figure 3.1. 7 Cemara

3.2 Collecting Data Method

The collection data methods of this research are consists of:

a. Observation

   The researchers do the observation directly in 7 Cemara get information about initial condition.

b. Interview

The researchers do interview to get information about 7 Cemara profile and get permission to acquire the data.

c. Data Collection

The researchers get the 100 receipts as the data for 5 days and directly filtering each receipts that fulfill requirement.

3.3 Types of Data

a. Primary Data

The primary data of this research is the data that is directly collected from the object of research which is 7 Cemara. The data includes 100 transaction receipts acquired from customers of 7 Cemara and the cashier of 7 Cemara, as well as the profile of 7 Cemara which is obtained by interviewing the workers of 7 Cemara.

b. Secondary Data

The secondary data used in this research is obtained from literature review about price discounting and association rules.

3.4 Flowchart

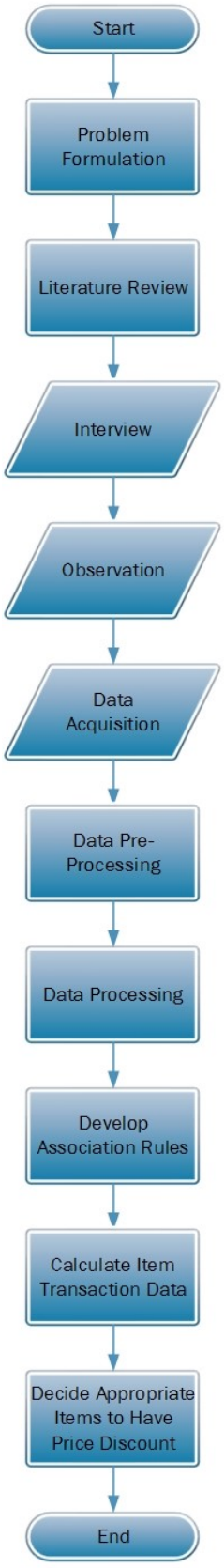

Figure 3. 2. Flowchart of Research

The figure above shows the flow of this research which includes the following procedures:

a. Problem Formulation

   The first procedure needed in this research is formulating the problem to overcome in this research which is price discounting of particular products based on AR-MBA in 7 Cemara minimarket.

b. Literature Review

   In order to find the solutions, literature review regarding to price discounting and AR-MBA is needed

c. Interview

   Interview is needed in order to find data regarding to the profile of research object, and initial condition of research object, as well as obtaining permission to do research

d. Observation

   Observation is a supporting need for initial condition data of research object and departments in the research object

e. Data Acquisition

   This is the phase where the transaction receipt is acquired. The transaction receipt is required from the customer and cashier of the research object

f. Data Pre-processing

   The data acquired with less than two item from different department is removed from the initial acquired data until the clean data obtained is 100 transaction data. After reduction and cleaning, the data is integrated and transformed into binary data which will be the input for *Rapid Miner* software

g. Data Processing

   In *Rapid Miner*, the binary data is placed as input which must pass through several processes which includes minimal support and confidence level deciding.

h. Develop Association Rule

   The output of *Rapid Miner* shows the relationship analysis between each department and the relationship with lifting index bigger 1.0 is selected for the association rule

i. Calculate Item Transaction Data

Using *Microsoft Excel* again, the department which are included in the association rules is calculated for each item purchase in the related department using *countif* formula

j. Decide Appropriate Items to Have Price Discount

Based on the previous procedure, the item with the lowest sales interest is given price discount to increase sales of that product

CHAPTER IV
RESULT AND CONCLUSION

4.1 Initial Condition of The Research Object

7cemara is such as minimarket with wifi@id wifi connection that located in Kaliurang Street Km 5.8 Kentungan. 7cemara established on January 2014. From first open until now, 7cemara always crowded by customer, the customer always buys some product in 7cemara but there are product or item oftenly bought and rarely bought. Then from first open until now, 7cemara not give discount to all products till now, so there is no discount in 7cemara until now. The figure below shows the current layout of 7cemara

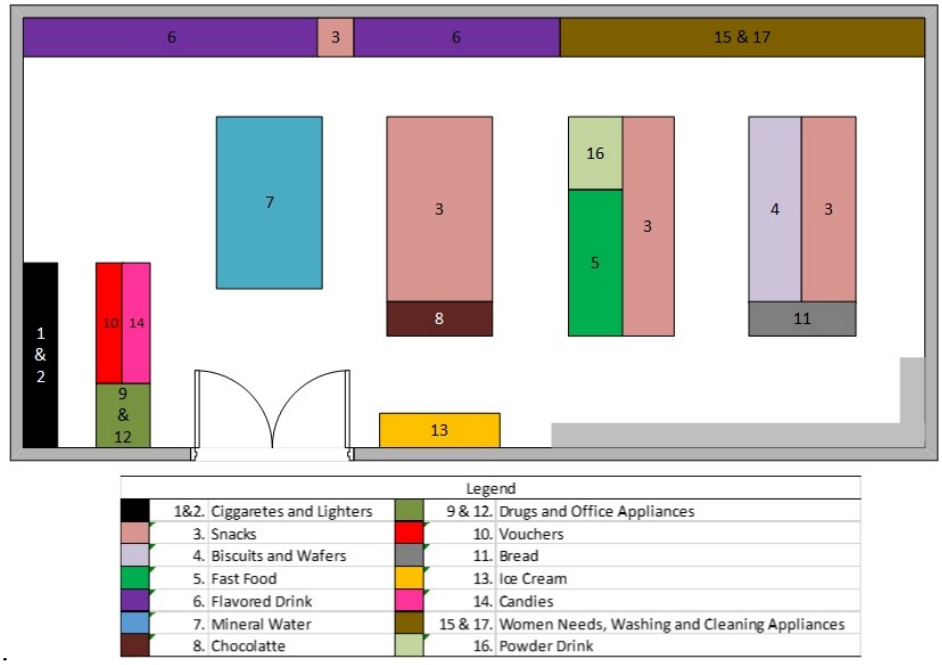

Figure 4. 1. Layout of 7 Cemara

4.2 Rapid Miner Output

No.	Premises	Conclusion	Support	Confide...	LaPla...	Gain	p-s	Lift	Convi...
7	Departemen 6	Departemen 1	0.180	0.316	0.752	-0.960	0.032	1.215	1.082
11	Departemen 1	Departemen 6	0.180	0.692	0.937	-0.340	0.032	1.215	1.398
10	Departemen 5	Departemen 6	0.050	0.625	0.972	-0.110	0.004	1.096	1.147
6	Departemen 6	Departemen 3	0.170	0.298	0.745	-0.970	0.010	1.065	1.026
9	Departemen 3	Departemen 6	0.170	0.607	0.914	-0.390	0.010	1.065	1.095
4	Departemen 7	Departemen 3	0.080	0.276	0.837	-0.500	-0.001	0.985	0.994
5	Departemen 3	Departemen 7	0.080	0.286	0.844	-0.480	-0.001	0.985	0.994
3	Departemen 6	Departemen 10	0.150	0.263	0.732	-0.990	-0.015	0.907	0.964
8	Departemen 10	Departemen 6	0.150	0.517	0.891	-0.430	-0.015	0.907	0.891

Figure 4. 2. Rapid Miner Output

The figure above shows the output of Rapid Miner software with minimum confidence level and minimum support level respectively 0.15 and 0.20.

4.3 Analysis of The Data Processing Result from The Point of View of Its Customer Behavior

From Rapid Miner Output Figure above that already analyze by calculating and classify receipts as the data input, we can see that output which already fulfill lift ratio requirement (lift ratio > 1) are output number 5, number 8, number 3 and number 7. There are rule for output which fulfill lift ratio requirement:

- Department 1 will bought with Department 6, with confidence level 69.2% and supported by 18% from overall data. Department 1 is cigarette and department 6 is beverage flavors.
- Department 5 will bought with Department 6, with confidence level 62.5% and supported by 5% from overall data. Department 5 is fast food and department 6 is beverage flavors.
- Department 3 will bought with Department 6, with confidence level 60.7% and supported by 17% from overall data. departemenet 3 is snacks and department 6 isbeverage flavors.

4.4 Solution Recommendations Based on The Associations Rule Result That You Found

After the researcher get the result and association rule from Rapid Miner software, then the researcher make solution recommendation based on the association rule above, from 100 transaction receipt also the researcher can find the product that oftenly bought and rarely bought by customer, there are some solution recommendation based on the association rule result:

- When the customer buy Dunhill Filter in department 1 and buy Vitamin Jambu in department 6, so cashier operator will give reduction price to Vitamin Jambu for Rp 2000,-.
- When the customer buy Pop Mie in department 5 and buy Lasegar in department in department 6, so cashier operator will give discount 5%.
- When the customer buy Taro in department 3 and Futami in department 6, so cashier operator will give discount 5%.

CHAPTER V
CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Therefore, the conclusion of this report are:

1. After the researcher do all calculation above, then the researcher got some associative relation from Rapid Miner software, there are:
   - Department 6 will bought with Department 1, with confidence level 31.6% and supported by 18% from overall data.
   - Department 5 will bought with Department 6, with confidence level 62.5% and supported by 5% from overall data.
   - Department 6 will bought with Department 3, with confidence level 29.8% and supported by 17% from overall data.
2. Then from associative relation above, the researcher can give solution recommendation about discount, and determine item that is often bought and rarely bought, there are:
   - When the customer buy Dunhill Filter in department 1 and buy Vitamin Jambu in department 6, so cashier operator will give reduction price to Vitamin Jambu for Rp 2000,-.
   - When the customer buy Pop Mie in department 5 and buy Lasegar in department in department 6, so cashier operator will give discount 5%.
   - When the customer buy Taro in department 3 and Futami in department 6, so cashier operator will give discount 5%.

5.2 Recommendation

After give recommendation about discount in 7 Cemara, the researcher thinks that maybe 7 Cemara should immediately implement the recommendation in order to get more customers, remembering that from first open until now, 7 Cemara not give discount in their product yet.

REFERENCE

Annie, L.C., & Kumar, A. 2012. Market basket analysis based on frequent item set mining. *International Journal of Computer Science Issues* **9**: 257-264

Hoanca, B. & Mock, K. 2011. Using market basket analysis to estimate potential revenue increases for a small university bookstore. *Conference for Information Systems Applied Research*. Wilmington.

Hoek, J. & Roelants, L. 1991. Some effects of price discounting on discounted and competing brands' sales. *Marketing Bulletin*.

Kaur, H. & Signh, K. 2013. Market basket analysis of sports store using association rules. *International Journal of Recent Trends in Electrical & Electronics Engineering* **3**: 81-85.

Kusumowidagdo, Astrid (2010) *Desain ritel komunikasikan strategi pemasaran bisnis dengan tepat*, PT Gramedia Pustaka Utama

ATTACHMENT

SEVEN CEMARA

NONGKRONG HEMAT

(Periode 1-31 Agustus 2015)

BELI DUNHILL FILTER DAN VIT JAMBU SEHARGA ~~16600~~ 14600

DAPATKAN POTONGAN HARGA 5%

*HANYA UNTUK KEDUA PRODUK DIATAS

DAPATKAN POTONGAN HARGA 5%

*HANYA UNTUK KEDUA PRODUK DIATAS

*syarat dan ketentuan berlaku