

ITEM ARRANGEMENT PATTERN IN WAREHOUSE USING APRIORI ALGORITHM (GIANT KAPASAN CASE STUDY)

¹RIFKI FAHRIAL ZAINAL, ²FARDANTO SETYATAMA

Department of Informatics Engineering, University of Bhayangkara Surabaya

Jalan Ahmad Yani 114, Surabaya 60231, Indonesia

e-mail: rifki@ubhara.ac.id, fardanto@ubhara.ac.id,

ABSTRACT

Giant is a retail company with supermarket format. Giant supermarkets should understand what are the items actually needed by their customers, particularly in easiness of choosing shop items. One of the method that can be used to analyze customer shopping behaviour pattern is analysis using the help of apriori algorithm. The analysis result, rules for item procurement are succesfully obtained. Rule that can be formed with minimum support and minimum confident highest values shows that the produced rule is {Sedap Mie Rasa Ayam Spc 69g → Cleo Air Minum Extra Oxygen 550 ml}. Based on the result, therefore Giant Kapasan should provide item Cleo Air Minum Extra Oxygen 550 ml when it sells Sedap Mie Rasa Ayam Spc 69g.

Keywords: Data Mining, Apriori Algorithm, Customer Shopping Behaviour Analysis.

1. Background

Business growth and competition in global trade through free market economy and advancing information technology bring higher competition and also make companies more open to satisfy customer demands. Companies have to implement good business strategy to be able to compete and keep their market. Competition within business could not be separated from information technology which becomes a trending topic at the moment. Giant is a retail company which sells its products through supermarkets. Giant always try to satisfy its customer by offering its quality products, excellent services trough friendly approach, and also creating fun shopping environment, but at the same time it is also competing with other supermarket. Therefore, strategies are needed to defend this retail business. Supermarkets should understand what are really needed by their customers, particularly in easiness of choosing shop items. For example the placement of shop items which are arranged in racks should be adapted from a customer's shopping pattern. Especially when they know their customer's pattern tendency in shopping activities, which usually buy things with a close product relation and they are bought together. So it is important to position shop items according to the customer's consumption pattern and actually it could affect customer's shopping desire as well as the number of product sales (Albion Research, 2007). A method which can be used to analyze customer's shopping behaviour pattern is Market Basket Analysis (MBA). This method is one of method in data mining which can be used to find frequently bought products from transaction data. MBA method uses apriori algorithm, an algorithm to provide association rule, with "if then" pattern.

2. Supporting Theory

2.1 Data Mining

Data mining is a process used to support decision making to find information patterns in a data. We can do this either by using query (in this case it is difficult to do) or by using an application which can do it automatically into a database called Discovery.

Discovery is a searching process in database to find hidden pattern without an idea or a hypothesis to start with. In other words, the application takes initiatives to find the pattern in a data without help from user to think about relevant question in first place. One of the pattern form that could be produced by data mining is association rule. Association Rule can be used to find relationship or causality.

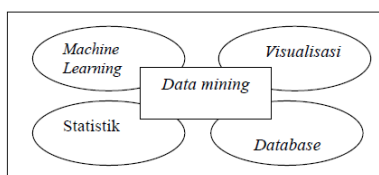


Figure 1. Data Mining has intersections with other disciplines

Machine Learning is an area in artificial intelligence theory that related to techniques development which can be programmed and learnt from existing data in the past. Pattern recognition, data mining and machine learning sometimes are used to call the same thing. These field intersect with probability theory and statistic and sometimes also with optimization. Machine learning becomes analysis tool in data mining. Figure 1 depicts the relationship of these fields (Santoso, 2007).

2.2 Apriori Algorithm

Apriori algorithm is the most famous algorithm to find high frequency pattern. High frequency pattern is a pattern of items in a database which have frequency or support beyond a threshold called minimum support. This high frequency pattern is used to develop associative rules and other data mining techniques.

Although there are more efficient algorithms are developed lately from Apriori, like FP-growth, LCM, etc., but apriori still becomes the most widely implemented algorithm in commercial product for data mining because it is considered as the most steadfast algorithm.

Apriori algorithm is divided into few steps called iteration or pass. Each of iteration produces high frequency pattern with the same length, starts from the first pass which produces high frequency pattern with length of one. At this first iteration, support from every item are counted by scanning the database. After the counts are finished, an item which has support higher than the minimum support is selected as the high frequency pattern with length 1 or often called 1 itemset. The abbreviation k-itemset means a set consist of k-item.

The second iteration produces 2-itemset, each of the set has 2 item. Firstly, 2-itemset candidate from combination of all 1-itemset is created. Then for each of this 2-itemset the supports are counted by scanning the database. Support, in this case, is the amount of transactions in the database which contain both item in the 2-itemset candidate. When all supports from all 2-itemset candidate are obtained, the 2-itemset candidate which fulfill the minimum support requirement can be assigned as 2-itemset which also becomes the high frequency pattern with length of 2.

For the next k iteration, it can be divided into few more parts:

1. Creating itemset candidate. K-itemset candidate are obtained from (k-1)-itemset combination from the previous iteration. One of the characteristic of apriori algorithm is the cut of k-itemset candidate which has a subset containing k-1 item, this candidate would not be included in the high frequency pattern with length k-1.
2. Counting support from each k-itemset candidate. Support from each k-itemset candidate are obtained by scanning database, the purpose is to count the amount of transaction which contain all items in that k-itemset candidate. This is also another characteristic of apriori algorithm where the counting are needed by scanning the whole database as many as the longest k-itemset.
3. Determining high frequency pattern. High frequency pattern containing k item or k-itemset is determined from k-itemset candidate which its support is greater than the minimum support.
4. When there is no more new high frequency pattern then the whole process is stopped. Otherwise, k is added by 1 and return to part 1.

Apriori rules are usually stated in form, as example:

{bread, butter} \rightarrow {milk}

(support = 0,4 confidence = 0,5)

This means, 0,5 from the whole transaction in database containing bread and butter item are also containing milk item. Meanwhile 0,4 from the whole transaction in database are containing the all three items.

It can also means when a customer buys bread and butter, he/she also has a 0,5 chance to buy milk. This rule is quite significant because it represents 0,4 of the whole transaction.

Associative analysis is defined as a process to find all association rules which qualify the minimum requirement for support (minimum support) and the minimum requirement for confidence (minimum confidence).

3. Design and Analysis

3.1 Data Flow Diagram

Picture 2 depicts the diagram to extract association rule at level 0, it is the most simple DFD. In the DFD, the extraction process of association rule needs 4 data inputs, which are transactional data, minsup (support minimum), mincof (confidence minimum) and item data. Transactional data is obtained from database, meanwhile minsup and mincof data are inputted by admin. Output data result from this process is association rule used to give the item sales pattern.



Figure 2 DFD Level 0

As seen in the figure, the k-item set frequent searching process needs two data inputs, transactional data and minimum support. Otherwise the data output from this process is all k-itemset frequent. This is the data that will be used as data input for the next process, called association rule forming process. So the process needs 2 data inputs, which are the frequent k-itemset (data output from k-itemset searching process) and the mincof data. The data produced from association rule forming process is all association rule data which qualify the minimum confidence.

Picture 3 depicts when k-itemset frequent searching process is divided into 3 subprocess:

1. Item Data Increment Subprocess
2. Transaction Data Increment Subprocess
3. Apriori Subprocess

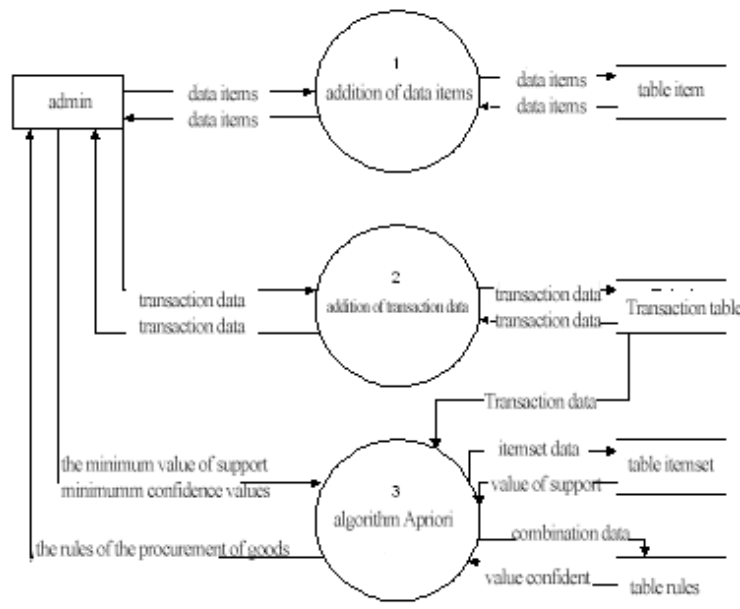


Figure 3 DFD Level 1

In the connection and query process, a database connection is created to connect to the database which contain the transactional data. From that connection, users can choose transactional table which will be

used as the data input. So, the data output from this process is transactional data retrieved by the query. This data will become the input for the next process.

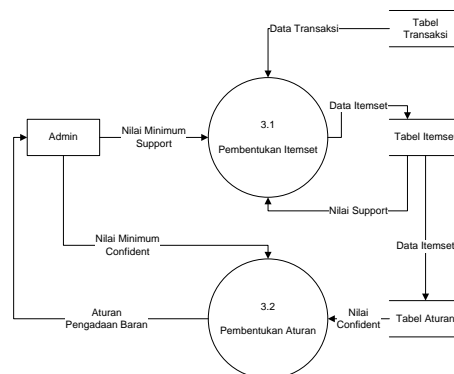


Figure 4 DFD Level 2 apriori

There are 2 data are needed in this process, they are transactional data retrieved by the query, and min-sup data inputted by admin. Data output result from this process is a frequent 1-item set as well as its value. Later, this data will be used for the last process. The last process is ($i > 1$) item set forming process. The data needed in this process is frequent 1-item set and also its tidlist. This process is done based on the min sup value which has been inputted before in the tidlist forming process. The data output produced is all the frequent item set which will be used in association rule forming process.

3.2 ERD

ERD (Entity Relational Diagram) is a model used to explain the relationship between data in database based on basic data objects which have relationship between relation. ERD used to model the data structure and the relationship between data, notations and symbols are used to depict them.

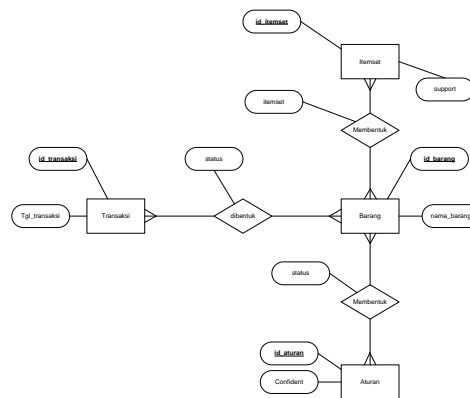


Figure 5 ERD design

4. Test Results

Tests are done on 92 data consisting 61 transactions in November 2013 and December 2013, and the data test consist of 31 transactions in January 2014. The tests are done by using variation in minimum support and minimum confident values.

Overall application test results with variation minimum support and minimum confident values are depicted in Table 1.

Table 1 Rule Forming Results

<i>MinSup</i>	<i>MinConf</i>	the number of rules on data 1	the number of rules on data 2	Error rate (%)
0,7	0,8	149	0	0
0,6	0,8	3833	1	100
0,5	0,8	7382	63	98,41
0,4	0,8	7075	708	58,05
0,3	0,8	6283	1897	32,16

5. Result Analysis

According to the test results, it can be concluded that the difference level between the first and second data is very high. This is caused by the shopping trend at Giant Kapasan every month.

6. Conclusion

1. A rule that can be formed with minimum support and minimum confident highest values using monthly data from November to December shows that the produced rule is {Sedap Mie Rasa Ayam Spc 69g → Cleo Air Minum Extra Oxygen 550 ml}
2. A rule that can be formed with minimum support and minimum confident highest values using monthly data in January shows that the produced rule is {Max Tea Lemon Tea 5x25g → Fruit Tea X-Tream Pet 500ml}
3. The difference between rule produced using November-Desember data and January data is 87,09%, this shows that item preparation pattern will be different for each month

7. Suggestion

1. Research could be improved by using more transactional data so there would be more accurate rules for item procurement.
2. Research could be improved by using more transactional data in another month so the item procurement pattern could be seen.
3. Research could be improved by using another data mining algorithm, like clusterization, so there could be additional information obtained.

REFERENCES

- [1] Santosa, Budi. 2007. Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta, Graha Ilmu.
- [2] Albion Research. "Data Mining : Market Basket Analysis." <http://www.albionresearch.com>. Albion Research Ltd., n.d. Web. 1 Jan 2017.
- [3] Han, Jiawei. 2005. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers Inc. San Fransisco, CA, USA.
- [4] Dongre, Jugendra, Prajapati, Lal Gend, & Tokekar, S.V. 2014. The role of Apriori algorithm for finding the association rules in Data mining. IEEE.