Grouping Goods With The Association Rule Method Using A Priori Algorithms In Modern Retail Stores

Arif Rahmadsyah[1], Nova Mayasari[2]*

[1,2] Faculty of Science and Technology, University of Pembangunan Panca Budi, Medan Indonesia
*Corresponding Author:
Email: maya7886@pancabudi.ac.id

Abstract.
The proliferation of retail companies in Indonesia makes competition quite fierce in this business. The availability of detailed information on products purchased from each customer transaction in the database will become data garbage, which continues to grow every day. Company managers will certainly be interested in knowing whether certain groups of products are consistently purchased together. They can use this data for optimal grouping and placement of products, and managers can also use this information for cross-selling, for promotions, for catalog design and to identify customer segments based on purchasing patterns. In this study, data mining with Association rule techniques was used to explore patterns of attribute relationships and frequent itemsets in the Retail database. The a priori paradigm is used to find large itemsets in the determination of association rules. The integration of association rules with the a priori paradigm has managed to find a number of patterns of relationships between attributes in retail databases.

Keywords: *Data mining, Association rule, AlgoritmaApriori, and Frequent Itemset.*

I. INTRODUCTION

Sales transactions that occur because sales obtain the approval of potential buyers to buy or use the products / services offered [1]. Sales transaction data owned by a retail company every day will increase in number [2][3][4]. Such large data can actually be a problem for companies if it cannot be utilized. The more data, the more companies need effort to sort out which data can be processed into information. If the data is left alone, then the data will only become meaningless waste for the companyIn this study, researchers will build an application that can group data on the purchase of goods based on their tendency to appear together in a transaction using an a priori algorithm. A priori algorithms belong to the type of association rules in data mining [5].

The rule that states the association between several attributes is often referred to as affinity analysis or market basket analysis [6][7].With the application of a priori algorithms in this study, it is hoped that a pattern will be found in the form of products that are often purchased together. This pattern can be used to place frequently purchased products together into an area close to each other, design the appearance of products in the catalog, design discount coupons (to be given to customers who buy a particular product), design package sales, and miscellaneous [8].From the background as exposed above, problems can be formulated in this study, namely: how to create an application that is able to group product data according to the level of tendency to appear together in a transaction purchase. The algorithm that will be applied to the application in this study is an a priori algorithm which is part of data mining.

II. METHODS

Data mining is defined as the process of finding meaningful new relationships, patterns and trends by sifting through very large amounts of data, stored in storage, using pattern recognition techniques such as statistical and mathematical techniques [8][9]. The relationship sought in Data mining can be in the form of a relationship between two or more in one dimension, for example in the product dimension, we can see the relationship between the purchase of a product and another product. In addition, relationships can also be seen between 2 or more attributes and 2 or more objects [10][11].

Association Mining

Association rule mining is a data mining technique to find associative rules between a combination of items. An example of an associative rule of purchasing analysis in a supermarket is that it is possible to know how likely a customer is to buy bread along with milk. With this knowledge, supermarket owners can arrange the placement of their goods or design marketing campaigns by using discount coupons for certain combinations of goods. As association analysis became famous for its application to analyze the contents of shopping carts in supermarkets, association analysis is also often referred to as market basket analysis [12]. Association analysis is also known as one of the Data mining techniques that are the basis of various other data mining techniques. In particular, one of the stages of association analysis called frequent pattern mining has attracted the attention of 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 the combination of these items in the database and confidence (certainty value), namely the strong relationship between items in the associative rule.

Apriori Algorithm

The issue of association rule mining consists of two sub-problems:

1. Find all combinations of items, called frequent itemsets, that have greater support than the minimum support.
2. Use frequent itemsets to generate the desired rules. For example, ABCD and AB are frequent, so the AB -> CD rule is obtained if the ratio of support (ABCD) to support (AB) is at least equal to the minimum confidence. This rule has a minimum of support because ABCD is frequent.

Apriori algorithms that aim to find frequent itemsets are run on a set of data. In the iteration to -k, all itemsets that have k items will be found, called k -itemsets. Each iteration contains two stages. For example, Oracle Data mining Fk represents the set of frequent k -itemsets, and Ck is the set of candidate k-itemsets (which have the potential to become frequent itemsets). The first stage is to generate candidates, where the set of all frequent (k-1) itemsets, Fk-1, found in the th iteration-(k-1), is used to generate candidate itemsets Ck. The candidate generate procedure ensures that Ck is a superset of the set of all frequent k-itemsets. The hash-tree data structure is used to store tsk. Then the data is scanned in the support calculation stage.For each transaction, candidates in Ck are populated into the transaction, defined using the hash-tree hashtree data structure and the support calculation value is raised. At the end of the second stage, the Ck value is tested to determine which of the candidates is frequent. The counter condition (terminate condition) of this algorithm is achieved when Fk or Tsk+1 is empty.

Pseudo-code:

Ck: *Candidate itemset of* size *k*
Lk: *frequent itemset of size k*

```
L1 = {frequent items};
for (k = 1; Lk! =0; k++) do begin
Ck+1 = candidates generated from Lk;
    for each transaction t in database do
        increment the count of all candidates in Ck+

        that are contained in t
        Lk+1  = candidates in Ck+1 with min_support
    end

return ∪k Lk;
```

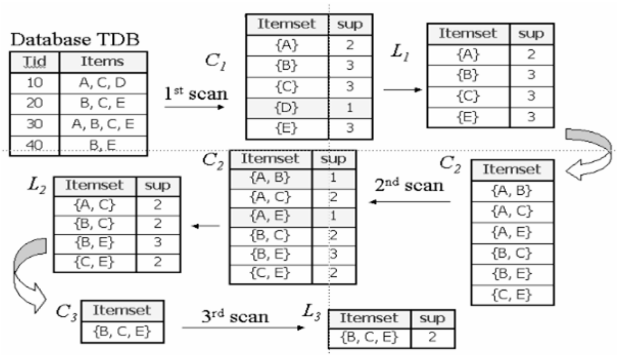

Fig 1. Illustration of Apriori Algorithms

From the illustration above it can be explained that if youproduce L_k, it is necessary to have a *candidate* k-itemset C_k formed from the process of joining L_{k-1} catatan convention: A priori assumes that the items in the transaction or *itemset* have been sorted according to the lexicographic order. The join process, $L_{k-1} \times L_{k-1}$, is carried out if ($_{k-2}$ *itemset* of L_{k-1} is "the same". For example, l_l and l_l are *itemsets* of L_{k-1}, in order for the *join* process to be carried out, it must be fulfilled: (l 1[1] = l 2[1]) ^ l 1[2] = l 2 [2]) ^...^ (l 1 [k-2] = l 2 [k-2]) ^ l 1 [k-1] < l 2 [k-1]). The condition (l 1 [k-1] < l 2 [k-1]) guarantees that there are no twins in the join process. So, the *itemset* resulting from the join process between *I*1 and *I*2 is l 1[1] l 1[2] ... l 1 [k-1] I2 [k-2]. The notation l 1 [j] states the jth item in l.

Stages of Association rules

Association analysis is also known as one of *the data mining* techniques that are the basis of various other *data mining* techniques. In particular, one of the stages of association analysis called *frequent pattern mining* has attracted the attention of many researchers to produce efficient algorithms. The basic methodology of association analysis is divided into two stages, namely (1) ahigh-frequency pattern nalisa**.** This stage looks for a combination of items that meet the minimum requirements of the support value in the database. The support value of an item is obtained by the following formula 1 [3]:

$$\text{Support (A)} = \frac{\text{Total Transactions contain A}}{\text{Total Transactions}}$$

while the support value of the 2 items is obtained from the following formula [3]:

$$Support\,(A, B) = P(A \cap B) = \frac{\text{Number of Transactions Containing A and B}}{\text{Total Transactions}}$$

Formation of Associative Rules

After all the high-frequency patterns are found, then an associative rule is sought that meets the minimum requirements for *confidence* by calculating the *confidence* of the associative rule A › B The *confidence* value of the rule A › B is obtained from the following formula:

$$Confidence = P\,(B|A) = \frac{\text{Number of Transactions Containing A and B}}{\text{Transaction Amount Contains A}}$$

III. DISCUSSION AND RESULTS

The number of rules generated provides many possibilities to see the patterns that appear in the Sales Transaction database. So as to provide various possibilities that can be used as a basis for making decisions. Not all of the rules found in this study were interpreted. What is interpreted is rules that have a high Lift value (objective reasons) and rules that have relevance to needs (subjective reasons) [13][14]. Lift is a ratio number that shows how many possibilities of finding an attribute (e.g., ID) appear along with other attributes (e.g., Qualification, Unique Number, and Certification mapel) compared to all occurrences of attributes being met Lift indicates a rule strength level over random events from antecedents and based on their respective support. This will provide information about improvements and probability increases of the consequent based on the antecedent. The elevator is defined as follows:

Lift* = *Confidence* / Expected *Confidence

where: *Confidence = (Number of Transactions have itemconsequent) / (Total number of transactions).* or other way:

$$Lift = \frac{\Pr(A \mid C)}{\Pr(C)}$$

When Lift is equal to 1 then A and B are independent because Pr(C|A) =Pr(C). When the probability of C occurring is affected by the occurrence of A greater than 1. Lift ratio provision is if the calculation result is below 1 then these items do not show any interrelationship between antacedent and consequent.

Suppose a list of transactions along with purchased items like table 1 below:

Table 1. Transaction List

Code Transaction	**Item Purchase Item**
001	SikatGigi,PastaGigi,SabunMandi,Detergen
002	SabunMandi,Bedak,Gula,Teh
003	Shampo,Bedak,SabunMandi
004	Bedak,SabunMandi,Detergen,Shampo,Softener
005	Shampo,Bedak,SikatGigi,PastaGigi,Softener,Detergen
006	Gula,Teh,SabunMandi,SikatGigi,PastaGigi
007	Detergen,SabunMandi,Softener,PastaGigi,SikatGigi
008	SabunMandi,Gula,Teh,Detergen
009	Softener,Detergen,Gula,Teh,SabunMandi
010	Detergen,Gula,Softener,Teh

Figure 2 below is a view of the results of running an application program for the implementation of a priori algorithms.

Fig 2.*Running and Result Frequent Item Set*

By input minimum support 50% and minimum confidence 50% data, frequent item set results are obtained as follows:

Minimum Support = 50%

Minimum Confidence = 50%

1 Item set (L1)

****

{Bedak} = 4/10 = 40,000% >Rejected<

{Detergen} = 7/10 = 70,000% <Accepted>

{Gula} = 5/10 = 50,000% <Accepted>

{PastaGigi} = 4/10 = 40,000% >Rejected<

{SabunMandi} = 8/10 = 80,000% <Accepted>

{Shampo} = 3/10 = 30,000% >Rejected<

{SikatGigi} = 4/10 = 40,000% >Rejected<

{Softener} = 5/10 = 50,000% < Accepted >

{Teh} = 5/10 = 50,000% < Accepted >

2 Item set (L2)
****
{Bedak,Detergen} = 2/10 = 20,000% >Rejected<
{Bedak,Gula} = 1/10 = 10,000% > Rejected <
{Bedak,PastaGigi} = 1/10 = 10,000% > Rejected <
{Bedak,SabunMandi} = 3/10 = 30,000% >Rejected<
{Detergen,SabunMandi} = 5/10 = 50,000% < Accepted >
{Detergen,Shampo} = 2/10 = 20,000% >Rejected<
{Detergen,SikatGigi} = 3/10 = 30,000% >Rejected<
{Detergen,Softener} = 5/10 = 50,000% < Accepted >
{Detergen,Teh} = 3/10 = 30,000% >Rejected<
{Gula,PastaGigi} = 1/10 = 10,000% >Rejectedk<
{Gula,SikatGigi} = 1/10 = 10,000% >Rejected<
{Gula,Softener} = 2/10 = 20,000% >Rejected<
{Gula,Teh} = 5/10 = 50,000% < Accepted >
{PastaGigi,SabunMandi} = 3/10 = 30,000% >Rejected<
3 Item set (L3)
****
{Bedak,Detergen,Gula} = 0/10 = ,000% > Rejected <
{Bedak,Detergen,PastaGigi} = 1/10 = 10,000% > Rejected <
{Bedak,Detergen,SabunMandi} = 1/10 = 10,000% > Rejected <
{Bedak,Detergen,Shampo} = 2/10 = 20,000% > Rejected <
{Bedak,Detergen,SikatGigi} = 1/10 = 10,000% > Rejected <
{Bedak,Detergen,Softener} = 2/10 = 20,000% > Rejected <
4 Item set (L4)
****
{Bedak,Detergen,Gula,PastaGigi} = 0/10 = ,000%> Rejected <
{Bedak,Detergen,Gula,SabunMandi} = 0/10 = ,000% > Rejected <
{Bedak,Detergen,Gula,Shampo} = 0/10 = ,000% > Rejected <
{Bedak,Detergen,Gula,SikatGigi} = 0/10 = ,000% > Rejected <
{Bedak,Detergen,Gula,Softener} = 0/10 = ,000% > Rejected <
{Bedak,Detergen,Gula,Teh} = 0/10 = ,000% > Rejected <

And so on unti l9 Item Set (L9)

List Interesting *Rules* :
****
Detergen->SabunMandi [50,00%,71,43%] < Accepted >
SabunMandi->Detergen [50,00%,62,50%] < Accepted >
Detergen->Softener [50,00%,71,43%] < Accepted >
Softener->Detergen [50,00%,100,00%] <Accepted>
Gula->Teh [50,00%,100,00%] <Accepted>
Teh->Gula [50,00%,100,00%] <Accepted>

IV. CONCLUSION

From the results of the previous explanation and description, it can be concluded that apriori algorithm has been successfully applied to see the rules of association between products on transactions that occur in the sales database at retail companies. To implement an a priori algorithm for finding the rules of product associations in the application as we built requires the adjustment of tables and columns from an existing database.The tendency of the pattern formed from the Association rule Mining from the experiment above is that there is a pattern if the customer buys Detergent, then he will also buy Bath Soap, as well as if the customer buys Detergent, then he will also buy Softener. Similarly, there is a pattern of buying customers who will buy Sugar as well as Tea. This is useful for determining the layout and grouping of goods to be sold in retail companies. So that the feeling of pleasure or disappointment of customers that arises after comparing the results of the layout of a product that is thought of with the expected wishes will lead to customer satisfaction [15].

V. THANK YOU

The author thanked the Faculty of Science and Technology, Universitas Pembangunan Panca Budi Medan for providing the author with the opportunity to contribute to scientific work as a final study report.

REFERENCE

[1] Bambang Suwarno (2019). Essential: Pengantar Manajemen 1 (Strategi Pemasaran Modern). CV. Andalan Bintang Ghonim. Medan.

[2] H. D. Anggraeni, R. Saputra, and B. Noranita, (2014). Aplikasi Data Mining Analisis Data Transaksi Penjualan Obat Menggunakan Algoritma Apriori (Studi Kasus di Apotek Setya Sehat Semarang), ***Journal of Informatics and Technology***, vol. 2, no. 2, pp. 24-30. https://ejournal3.undip.ac.id/index.php/joint/article/view/6280

[3] Andriani, Anik. (2013). Aplikasi Data Mining Market Basket Analysis Penjualan Suku Cadang Sepeda Motor Menggunakan Metode Association Rules Pada PT. Sejahtera Motor Gemilang. ***Jurnal Ilmiah Tekhnik Informatika*** UniversitasNusantara PGRI.Kediri. http://simki.unpkediri.ac.id/detail/10.1.03.02.0170

[4] Sijabat, Alimancon. (2015). Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Tree. ***Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI)***. Vol. 5 No 3, pp.7-12.

[5] Abidi dan Yu-N, (2000), A Convergence of Knowledge Management and *Data mining*: Towards 'Knowledge-Driven' Strategic Services. ***3rd International Conference on the Practical Applications of Knowledge Management*** (PAKeM'2000), April 12-14 2000, Manchester.

[6] R. Karthiyayini & R. Balasubramanian (2016). Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis. ***International Journal of Advanced Research in Computer Science and Software Engineering.*** Volume 6, Issue 10, pp. 241-246.

[7] Manpreet, Kaur & Shivani, Kang (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. ***International Conference on Computational Modeling and Security. Elsevier.*** Procedia Computer Science 85 (2016) 78 – 85. https://doi.org/10.1016/j.procs.2016.05.180

[8] Berry, Michael J.A dan Linoff, Gordon S., (2004) ***Data mining Techniques for Marketing, Sales, Customer Relationship Management***. Second Editon, Wiley Publishing, Inc.

[9] Larose, Daniel T, (2005) ***Discovering Knowledge in Data: An Introduction to Data mining***, John Willey & Sons. Inc

[10] Ponniah, P. (2001). ***Data Warehouse Fundamentals: A Comprehensive Guide for IT Professional,*** John Willey & Sons. Inc

[11] Pramudiono, I., (2006), ***Apa itu data mining***? http://data*mining*.japati.net/cgi-bin/indodm.cgi?bacaarsip&1155527614&artikel,

[12] Srikant, R., Vu, Q., dan Agrawal R., (1997), **Mining Association rules with Item Constraints**, www.aaai.org

[13] Turban, E., dkk, (2005), ***Decicion Support Systems and IntelligentSystems***, Andi Offset.

[14] JE Aronson, TP Liang, RV MacCarthy (2005). ***Decision Support Systems and Intelligent Systems***. International Edition, Seventh Edition.

[15] Suwarno, B. (2022). An Analysis of Purchase Decisions on Customer Satisfaction Through Customer-Based Brand Equity and Product Innovation: Concequences for Air Conditioner Panasonic in Medan. ***International Journal of Science, Technology & Management***, 3(3), 605-620. https://doi.org/10.46729/ijstm.v3i3.515.