cam0

by Cam Cam

Consumer Customs Analysis Using the Association Rule and Apriori Algorithm for Determining Sales Strategies in Retail Central

Ahmad Heru Mujianto[1,*], *Chamdan* Mashuri[1], *Anita* Andriani[2], and *Febriana Dwi* Jayanti[3]

[1]Information System Department of Information Technology Faculty, Hasyim Asy'ari University, Jombang – Indonesia
[2]Information Management Department of Information Technology Faculty, Hasyim Asy'ari University, Jombang – Indonesia
[3]Information Department of Information Technology Faculty, Hasyim Asy'ari University, Jombang – Indonesia

Abstract. The sustainability of a company will not be separated from the role of consumers in conducting transactions. In fact, a consumer has different behaviour and character, therefore as a company owner must be able to analyze the patterns or habits of consumers in making transactions. This also happens in the retail center X, which has problems in the sales process, such as products running out of stock and unsold products and the most popular products and products that are not in demand by consumers. Therefore we need an analysis of consumer habits in conducting transactions. The method of association rule with Apriori algorithm is able to be applied well in the analysis of the habits of consumer transactions in the central retail X. The results of the calculation obtained an average percentage of the value of support 33%-40% and the value of confidence 43%-80%. The results of applying the association rule method with Apriori algorithm can help recommend central retail X owners in structuring product and determine strategic steps in increasing sales, such as providing discounts or promos for certain products.

Keywords: **Analysis; Consumer; Association Rule; Apriori; Retail.**

1 Introduction

Consumers are a very influential aspect in business continuity, where consumers are everyone involved in buying and selling activities to meet their needs. The consumer behaviour varies, which can be seen based on three stages, namely when before making a purchase, at the time of purchase, and after purchase. Therefore it is important for businesses to understand every consumer behaviour to determine the right marketing strategy for business continuity [1]. To understand consumer behaviour, analysis is needed on a collection of consumer transaction data, to dig deeper information based on these data or known as data mining.

Data mining is a data processing technique by extracting and identifying large data databases to extract useful information and increase user knowledge. Whereas associations rule is one method in data mining, this method is often used in analysis consumer spending patterns, to see the interrelationship between goods from transactions carried out by customers [2].

The algorithm used in the most popular association rule is Apriori algorithm, the algorithm is often used in analysis transaction data or basketball market. The application of Apriori algorithm is used for data retrieval by paying attention to associative rules to determine associative relationships of a combination. The procedure in determining associative rules starts with

collecting transaction data that is carried out by consumers from time to time to find out what items are purchased by consumers so that it will be known how much the relationship between goods and can be implemented into an associative rule. The application of Apriori algorithms can be done on transactions that occur in stores, supermarkets and retail [3].

Previous research has been conducted by Gull and Pervaiz, who analyzed consumer behaviour when shopping online at an online shopping site with the method used is the association rule. The results of the study obtained data about the number of customers who visited the site and made a purchase transaction. As well as the combination and interrelation between goods purchased by consumers [4].

In this paper applying Apriori algorithm in analyzing consumer purchasing habits carried out at the retail center X, where at the center of retail X there were problems related to sales, which included the existence of unsold goods but some products ran out of inventory. On the other hand, there are many consumers with different habits and characters. Another problem occurs in the retail center experiencing goods arrangement that has not attracted consumers or can be attributed to the absence of a sales strategy that is applied to attract consumers [5]. Out of stock and goods that are less desirable are important and emergency problems in the world of goods trading business, because it has a

---

profound effect on the profits received by the company. These problems need a formula to overcome some of the phenomena that occur. Therefore, a sales strategy is needed that can analyze sales transactions so that it can provide information regarding less salable goods and goods that are out of stock. In this study, the associational rule of the Apriori algorithm was used to analyze consumer habits in retail X to determine the sales strategy. The application of association rules with Apriori algorithm will know the relationship and possible combination of items purchased by consumers at the retail center X [6]. The final results of this paper are expected to provide recommendations on what items should be increased, what items should be placed side by side and what items are less attractive to sellers.

2 Methodology

2.1. Association Rule

Association rule is one technique in data mining to determine the relationship between items in a dataset (a set of data) that has been determined. This concept itself is derived from the terminology of market basket analysis, namely the search for relationships of several products in a purchase transaction. This technique looks for possible frequent combinations of an item set [6].

An example of association rule is based on consumer purchase data at a retail store, consumer A {bread, milk, butter}, consumer B {milk, bread, eggs} and consumer C {bread, milk, jam}. Based on the three consumer transactions, it will produce an association rule, which is a link between one product and another product, that is, if you buy bread, you will buy milk [7]. The relationship of the association rules formed is denoted in $X \rightarrow Y$ numbers, where X and Y are a disjoint item set $(X \cap Y) = \emptyset$. In the association rule, we can calculate support and confidence [8]

Support is how much an item affects the overall transaction. Suppose that calculating support items A and B from all transactions (S), while the equation can be denoted as follows:

$$Support\ (A \rightarrow B \mid S) = \frac{n(S \cap A \cap B)}{n(S)} \ x\ 100\% \quad (1)$$

While confidence is a relationship between 2 items conditionally, for example how often consumers will buy items B will also buy item A. The following formula is used to calculate the value of confidence:

$$Confidence = (A \rightarrow B \mid S) = \frac{n(S \cap A \cap B)}{n(S \cap A)} \ x100\% \quad (2)$$

There are several algorithms that have been developed regarding association rules, but there is one classic algorithm that is often used, namely Apriori algorithm. The basic idea of this algorithm is to develop frequent item set. By using one item and recursively developing frequent item set with two items, three items and so on to frequent item set with all sizes.

The Apriori algorithm is an algorithm introduced by R.Agrawal and R.rikant in 1994. A priori algorithms aim to find frequent item sets that are run on a set of data. In the first iteration, all item sets that have k-items will be found, called k-item set. The main characteristic of Apriori algorithms is that all subset of frequent item sets are also applying frequent item sets [9]. The principles of Apriori algorithm are:

1) Collect a number of single items, get large items
2) Get a candidate pairs, count → large pairs of items
3) Get candidate triplets, count → large triplets from items and so on
4) For instructions: each subset of a frequent item set must be frequent

The two main processes in Apriori algorithm are steps that will be taken to obtain frequent item sets. Although Apriori algorithm is easy to understand and implement compared to other algorithms that are applied to the process of association rules. Apriori algorithms have a disadvantage, namely to search frequent item set, the priori algorithm must scan the database repeatedly for each item combination. This causes the amount of time needed to scan the database. In addition, it requires generating large candidates to get item combinations from the database [10].

Apriori algorithms are used to find association rules that meet the value limits of support and confidence. The a priori work method is to analysis a collection of items purchased together on several transactions. Calculation of a collection of data patterns that appear in the database through several iterations or repetitions. Iteration i calculates all data sets i (data set containing element i) that often appears. Each iteration consists of two steps, namely a candidate generation and candidate counting and selection (selection and calculation of candidates) [9].

One stage of the association analysis that attracted the attention of many researchers to produce efficient algorithms is the analysis of high frequency patterns. The importance of an association can be known by two benchmarks, namely: support and confidence. Support (supporting value) is the percentage combination of items in the database, while confidence (the value of certainty) is the strong relationship between items in the rules of association [7].

3 Result and Discussion

3.1 Implementation of Apriori Algorithm

This research was carried out and retrieved data in the central retail X, where in the retail center there were problems in product sales, so researchers wanted to analyze the habits of consumers when shopping at central retail X, to find out patterns and habits of consumers. The following is the transaction data that occurs in the central retail X shown in Table 1.

Table 1. Sales Transaction Data

No	Transaction	Product
1	TR001	Sugar, Coffee, Soap
2	TR002	Sugar, Bread, Toothpaste
3	TR003	Toothpaste, Soap, Coffee, Sugar
4	TR004	Sugar, milk
5	TR005	Soap, Bread
6	TR006	Bread, Milk, Sugar, Coffee
7	TR007	Shampoo, Soap, Bread, Sugar
8	TR008	Coffee, Soap, Toothpaste
9	TR009	Soap, Toothpaste, Bread
10	TR010	Shampoo, Toothpaste, Sugar, Coffee, Soap

To facilitate the calculation, then the product data in table 1 is coded and separated by each product as shown in Table 2 below.

Table 2. Item Data

Item	Code
Sugar	A
Coffee	B
Bread	C
Milk	D
Toothpaste	E
Soap	F
Shampoo	G

Based on data in Table 2, transaction data for each item is then made, as shown in Table 3.

Table 3. Item Transaction Data

Transaction	A	B	C	D	E	F	G
1	1	1	1	0	0	0	0
2	1	0	1	0	1	0	0
3	1	1	0	0	1	1	0
4	1	0	0	1	0	0	0
5	0	0	1	0	0	1	0
6	1	1	1	1	0	0	0
7	1	0	1	0	0	1	1
8	0	1	0	0	1	1	0
9	0	0	1	0	1	1	0
10	1	1	0	0	1	1	1
Σ	7	5	6	2	5	6	2

The next stage determines the value of Φ, which is given the value $\Phi = 3$, then the frequency value of item set from Table 3, there are 5 items greater than the value of Φ, so F1 = {{A},{B},{C},{E},{F}}. For a combination of 2 item set, a Table is needed to combine each item set. The following is a Table of 2 item set combinations that can be seen in Table 4.

Table 4. Combination of 2 item set

T	A	B	f	T	A	C	f	T	A	E	f
1	1	1	P	1	1	1	P	1	1	0	S
2	1	0	S	2	1	1	P	2	1	1	P
3	1	1	P	3	1	0	S	3	1	1	P
4	1	0	S	4	1	0	S	4	1	0	S
5	0	0	S	5	0	1	S	5	0	0	S
6	1	1	P	6	1	1	P	6	1	0	S
7	1	0	S	7	1	1	P	7	1	0	S
8	0	1	S	8	0	0	S	8	0	1	S
9	0	0	S	9	0	1	S	9	0	1	S
10	1	1	P	10	1	0	S	10	1	1	P
Σ			4	Σ			4	Σ			3

T	A	F	f	T	B	C	f	T	B	E	f
1	1	0	S	1	1	1	P	1	1	0	S
2	1	0	S	2	0	1	S	2	0	1	S
3	1	1	P	3	1	0	S	3	1	1	P
4	1	0	S	4	0	0	S	4	0	0	S
5	0	1	S	5	0	1	S	5	0	0	S
6	1	0	S	6	1	1	P	6	1	0	S
7	1	1	P	7	0	1	S	7	0	0	S
8	0	1	S	8	1	0	S	8	1	1	P
9	0	1	S	9	0	1	S	9	0	1	S
10	1	1	P	10	1	0	S	10	1	1	P
Σ			3	Σ			2	Σ			3

T	B	F	f	T	C	E	f	T	C	F	f
1	1	0	S	1	1	0	S	1	1	0	S
2	0	0	S	2	1	1	P	2	1	0	S
3	1	1	P	3	0	1	S	3	0	1	S
4	0	0	S	4	0	0	S	4	0	0	S
5	0	1	S	5	1	0	S	5	1	1	P
6	1	0	S	6	1	0	S	6	1	0	S
7	0	1	S	7	1	0	S	7	1	1	P
8	1	1	P	8	0	1	S	8	0	1	S
9	0	1	S	9	1	1	P	9	1	1	P
10	1	1	P	10	0	1	S	10	0	1	S
Σ			3	Σ			2	Σ			3

In Table 4, there is a *S* code which means 2 items are not sold together or nothing is sold, as *P* code indicates that the two items are sold together. The frequency of set items from Table 4 is as follows: F2 = {{A,B}, {A,C}, {A,E}, {A,F}, {B,E}, {B,F}, {C,F}, {E,F}}. The combination of 2 set items is continued for a combination of 3 set items, shown in Table 5.

Table 5. Combination of 3 item set

T	A	B	C	f	T	A	B	E	f
1	1	1	1	P	1	1	1	0	S
2	1	0	1	S	2	1	0	1	S
3	1	1	0	S	3	1	1	1	P
4	1	0	0	S	4	1	0	0	S
5	0	0	1	S	5	0	0	0	S
6	1	1	1	P	6	1	1	0	S
7	1	0	1	S	7	1	0	0	S
8	0	1	0	S	8	0	1	1	S
9	0	0	1	S	9	0	0	1	S
10	1	1	0	S	10	1	1	1	P
Σ				2	Σ				2

T	A	B	C	f	T	A	B	E	f
1	1	1	0	S	1	1	1	0	S
2	1	0	0	S	2	1	1	1	P
3	1	1	1	P	3	1	0	1	S

4	1	0	0	S	4	1	0	0	S
5	0	0	1	S	5	0	1	0	S
6	1	1	0	S	6	1	1	0	S
7	1	0	1	S	7	1	1	0	S
T	A	B	C	f	T	A	B	E	f
8	0	1	1	S	8	0	0	1	S
9	0	0	1	S	9	0	1	1	S
10	1	1	1	P	10	1	0	1	S
∑			2		∑				1
T	A	B	C	f	T	A	B	E	f
1	1	1	0	S	1	1	0	0	S
2	1	1	0	S	2	1	1	0	S
3	1	0	1	S	3	1	1	1	P
4	1	0	0	S	4	1	0	0	S
5	0	1	1	S	5	0	0	1	S
6	1	1	0	S	6	1	0	0	S
7	1	1	1	P	7	1	0	1	S
8	0	0	1	S	8	0	1	1	S
9	0	1	1	S	9	0	1	1	S
10	1	0	1	S	10	1	1	1	P
∑		1			∑			2	

Based on the data in Table 5, the value of F3 = {}, because there is no number of item set frequency ≥ 2. Then determine the rule to be used based on F2 = {{A, B}, {A, C}, {A, E}, {A, F}, {B, E}, {B, F}, {C, F}, {E, F}}. The following are the rules formed which are shown in Table 6.

Table 6. Rule Association Candidates

No	If antecedent the consequent
1	If buy A then buy B
2	If buy B then buy A
3	If buy A then buy C
4	If buy C then buy A
5	If buy A then buy E
6	If buy E then buy A
7	If buy A then buy F
8	If buy F then buy A
9	If buy B then buy E
10	If buy E then buy B
11	If buy B then buy F
12	If buy F then buy B
13	If buy C then buy F
14	If buy F then buy C
15	If buy E then buy F
16	If buy F then buy E

3.2 Implementation of Association Rule

After getting the candidate rule value, then the value of support and confidence is calculated, using the formula:

$$Support\ (A \rightarrow B \mid S) = \frac{n(S \cap A \cap B)}{n(S)} \times 100\%$$

$$Confidence = (A \rightarrow B \mid S) = \frac{n(S \cap A \cap B)}{n(S \cap A)} \times 100\%$$

Here are the calculation results are shown in Table 7.

Table 7. Rule Association Calculation

If antecedent then consequent	Support	Confidence
If buy A then buy B	(4/10)x100% = 40%	(4/7)x100% = 57,1%
If buy B then buy A	(4/10)x100% = 40%	(4/5)x100% = 80%
If buy A then buy C	(4/10)x100% = 40%	(4/7)x100% = 57,1%
If buy C then buy A	(4/10)x100% = 40%	(4/6)x100% = 66,7%
If buy A then buy E	(3/10)x100% = 33,3%	(3/7)x100% = 42,9%
If buy E then buy A	(3/10)x100% = 33,3%	(3/5)x100% = 60%
If buy A then buy F	(3/10)x100% = 33,3%	(3/7)x100% = 42,9%
If buy F then buy A	(3/10)x100% = 33,3%	(3/6)x100% = 50%
If buy B then buy E	(3/10)x100% = 33,3%	(3/5)x100% = 60%
If buy E then buy B	(3/10)x100% = 33,3%	(3/5)x100% = 60%
If buy B then buy F	(3/10)x100% = 33,3%	(3/5)x100% = 60%
If buy F then buy B	(3/10)x100% = 33,3%	(3/6)x100% = 50%
If buy C then buy F	(3/10)x100% = 33,3%	(3/6)x100% = 50%
If buy F then buy C	(3/10)x100% = 33,3%	(3/6)x100% = 50%
If buy E then buy F	(4/10)x100% = 40%	(4/5)x100% = 80%
If buy F then buy E	(4/10)x100% = 40%	(4/6)x100% = 66,7%

Based on Table 7, the highest value of confidence is sought for the next multiplied by the value of support and confidence.

Table 8. Calculation of Support and Confidence

If antecedent then consequent	Support	Confidence	Support X Confidence
If buy B then buy A	(4/10)x100% = 40%	(4/5)x100% = 80%	0,128
If buy E then buy F	(4/10)x100% = 40%	(4/5)x100% = 80%	0,128

The calculation results in tables 7 and 8 show that the rules used are with the highest confidence values, namely

- If you buy B, you will buy A
  (If you buy coffee, you will buy sugar too)
- If you buy E, you will buy F
  (If you buy toothpaste, you will buy soap too)

These rules will be implemented by retail owners to regulate the layout of these products based on patterns or habits of consumers in buying products.

3.3 Implementation of Sales Strategy

Previously presented samples of sales transaction data from 7 retail products X in Table 1, the following graphs of sales are based on product items.
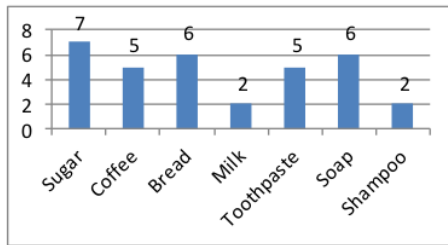
Fig. 1. Product Sales Graph

The data is sales transaction data for one week. Based on the graph in figure 1, it can be seen that the sugar products are most popular with consumers, while the milk products and shampoo with the least sales.

Based on the results of the association rule calculation and product sales data, thus the seller can take strategic steps to increase sales, which include:

1) Increasing the stock of sugar so that there is no shortage or out of stock of products, while the stock of milk and shampoo can be reduced so that there is no accumulation of stock which results in product expiration and is not sold.

2) Adjust the layout of the most desirable and interrelated products (combinations), so as to facilitate consumers to find the products they are looking for, such as placing sugar in addition to coffee.

3) Give discounts on products that are less desirable, such as discounts for purchasing milk and shampoo.

4) Providing promos to every combination product buyer, such as promotions for every purchase of soap and toothpaste.

By implementing the strategic steps above, it is expected that retail X can increase sales and revenues and be able to compete with other retailers.

4 Conclusion

Based on the results of the discussion above, researchers can draw some conclusions that include, the results of calculations with association rules and a priori algorithms are able to analyze consumer habits in making purchases, with a percentage average value of support 33% - 40% and confidence values 43% - 80%. Based on the results of these calculations can be used as a recommendation for retail owners to arrange the arrangement of product catalogs and take strategic steps to improve product marketing. The association rule method and a priori algorithm are able to be well applied in analyzing consumer spending patterns. However the Apriori algorithm still has weaknesses which have to scan the entire database as iterations are formed so that the time needed will increase.

References

1. Setiadi, *Consumer behavior*, Kencana (2015)

2. D. Prajapati, S. Garg, N.C. Chauhan, *Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment*, Future Computing and Informatics Journal **2**, 19-30 (2017)

3. A. Bhandari, et al, *Improvised apriori algorithm using frequent pattern tree for real time applications in data mining*, Procedia Computer Science **46**, 644 – 651 (2015)

4. M. Gull, A. Pervaiz, *Customer Behavior Analysis Towards Online Shopping using Data Mining*, International Multi-Topic ICT Conference **5** (2018)

5. S. Yamada, et al, *Visualization of relations of stores by using Association Rule Mining*, ICT and Knowledge Engineering, 11-15 (2015)

6. Witten, et al, *Data Mining. Partical Machinelearning Tools and Techniques 2nd Education*, Morgan Kaufmann Publisher (2005)

7. Kusrini, E.T. Luthfi, *Algorithm Data Mining*, Andi (2009)

8. H. Hammami, et al, *Classification and Projection of Spatial Association Rules*, SETIT (2012)

9. R. Sumithra, S. Paul, *Using distributed apriori association rule and classical apriori mining algorithms for grid based knowledge discovery*, ICCNT (2010)

10. J. Du, et al, *Research and Improvement of Apriori Algorithm*, ICIST, 117-121 (2016)

cam0

PRIMARY SOURCES

1	prosiding.seminar-id.com Internet Source	5%
2	informatikalogi.com Internet Source	3%
3	www.coursehero.com Internet Source	1%
4	seaninstitute.org Internet Source	1%
5	Chamdan Mashuri, Ahmad Heru Mujianto, Hadi Sucipto, Rinaldo Yudianto Arsam, Ginanjar Setyo Permadi. "Production Time Optimization using Campbell Dudek Smith (CDS) Algorithm for Production Scheduling", E3S Web of Conferences, 2019 Publication	1%
6	Jyx.jyu.fi Internet Source	1%
7	ejournal.raharja.ac.id Internet Source	1%

8   ojs.unud.ac.id
Internet Source
   1 %

9   Lisna Zahrotun, Dewi Soyusiawaty, Rahma Sara Pattihua. "The Implementation of Data Mining for Association Patterns Determination Using Temporal Association Methods in Medicine Data", 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2018
Publication
   1 %

10   Suprianto Panjaitan, Sulindawaty, Muhammad Amin, Sri Lindawati, Ronal Watrianthos, Hengki Tamando Sihotang, Bosker Sinaga. "Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns", Journal of Physics: Conference Series, 2019
Publication
   1 %

11   eprints.undip.ac.id
Internet Source
   1 %

12   en.wikipedia.org
Internet Source
   <1 %

13   pdffox.com
Internet Source
   <1 %

14   "ICCOEE2020", Springer Science and Business Media LLC, 2021
Publication
   <1 %

15 W Septiani, I A Marie, D Sugiarto, L Hakim. "The Pattern Failure Analysis of Sulfuric Acid Production Process with the Association Rules Algorithm Apriori", IOP Conference Series: Materials Science and Engineering, 2019
Publication

<1 %

16 www.ijais.org
Internet Source

<1 %

17 "Proceedings of First International Conference on Mathematical Modeling and Computational Science", Springer Science and Business Media LLC, 2021
Publication

<1 %

18 Dedy Rahman Prehanto, Aries Dwi Indriyanti, Chamdan Mashuri, Ginanjar Setyo Permadi. "Soil Moisture Prediction using Fuzzy Time Series and Moisture sensor Technology on Shallot Farming", E3S Web of Conferences, 2019
Publication

<1 %

19 Y Sutisnawati, M Reski. "Looking for Transaction Data Pattern Using Apriori Algorithm with Association Rule Method", IOP Conference Series: Materials Science and Engineering, 2019
Publication

<1 %

| Exclude quotes | On | Exclude matches | Off |
| Exclude bibliography | On | | |