
Sales Transaction Data Analysis using Apriori Algorithm to Determine the Layout of the Goods

Taqwa Hariguna ^a, Uswatun Hasanah ^a, Nindi Nofi Susanti ^{a, *},

^aSystem Information Study Programme, Amikom Purwokerto University, Indonesia

* Corresponding author

Abstract

In a shop, usually apply a sales strategy in order. The sales strategy can be in the form of determining the layout of goods so that they are close to one another. Determining the layout of items can be based on items that are often purchased simultaneously. Searching for items that are often purchased together can be done using data mining techniques, which is processing data to become more useful information. Sales transaction data processing can be done using apriori algorithm. Apriori algorithm is the most famous algorithm for finding high-frequency patterns and generating association rules. From the results of the discussion and data analysis, there were 3 (three) association rules formed, namely "If you buy Milo Active 18 grm, then buy ABC Kopi Susu 31G" with support 0.36% and 75% confidence, "If you buy Dancow 1 + Honey 200 grm, then buy Ice Cream Corneto" with support 0.36% and confidence 60%, "If you buy SIIP Roasted 6.5 grm, then buy Davos Strong 10 grm" with support 0.36% and 75% confidence. From the association's rules can be used as decision making to determine the layout of goods that are likely to be purchased simultaneously by the buyer.

Keywords: algorithm; Apriori; Association Rule

1. Introduction

In the business world of minimarket, data is vital information to expand the scope of business. To achieve this, there are several ways to be done namely improve product quality, ensure stock inventory and determine layout of goods. All these ways can be done by analyzing minimarket data. In the business world, such as a shop, minimarket, and supermarket is required to develop its business strategy in selling products [1][2]. One of the sales strategies that are done is to determine the layout of items based on the items that are often purchased. Pandak Baturraden Minimarket is a place that sells a variety of daily necessities. In order to increase the convenience of minimarket and service quality, it is necessary to make decision making to determine the Minimarket sales strategy by determining the layout of the goods to the adjacent distance for the goods related to one with the others [3][4].

Based on the interview with the Minimarket owner currently minimarket Pandak Baturraden has not a sales strategy yet such as layout determination of goods based on the goods that are often purchased. At minimarket Pandak Baturraden, the placement of the layout of the goods still not neatly arranged, the placement of goods is still very random so that there is no association of goods with the other. The placement of goods/products is still based on the price of products that have a high selling price placed behind the cashier to avoid the risk of loss. Then the product that type of rotation is quickly placed on the front shelf. Product placement shelf has several types namely TPE 29, 49, and 56; the meaning is on one shelf; there are 29, 49, and 56 sampling products. The products on the shelves are placed based on similar categories.

Since the Minimarket Pandak Baturraden only applies the layout of goods/products based on the price and category only, the Minimarket is challenging to determine which goods/products are often bought by consumers. Based on the results of the Minimarket, the interview experienced a decrease in the number of transactions that usually in February 2018 as much as 200 transactions per day to 100 transactions only per day in March 2018. Therefore, Pandak Minimarket requires a new sales strategy that is the determination of the layout of goods based on the items that are often purchased. To determine which frequently purchased items can use the sales transaction data on a Pandak minimarket. One of the techniques used to process transaction data is through data mining.

Data mining is a series of processes to explore the value of a data collection of knowledge that has not been known manually [5]. Data in question can be a sales transaction data a minimarket that was used only as a sales report,

inventory control and so on [6][7]. With the technique of data mining to process, several transaction data can be used in deciding to obtain more useful information than ever before. To obtain information from a data mining technique should be used, several methods contained in data mining; one such method is the method of association. Also, in data mining, several algorithms are often used to analyze shopping carts using sales transaction data such as a priori algorithm, FP-Growth algorithm, and Linear Congruent Method (LCM) algorithm[8][11].

In this research, researchers use a priori algorithm as a method of data processing of sales transactions at Minimarket because the priori algorithm will produce a combination pattern of items and rules as knowledge and information of the transaction data [9]. Using a priori algorithm in processing the transparency data can be found as an association rule that is a data mining technique to find the associative rules of a combination of items with the "if-then" pattern. The search process of the association rules starts with the transaction data processing, and then search the relationship between the purchased products.

Information obtained after using apriori algorithm can be used as a decision making in the arrangement of the product layout in the Minimarket. With the arrangement of product layout that has the relevance of one and the other, can facilitate the customer in the process of searching the desired product or item. In a minimarket, employees are still ignoring the problem of product layout so that they are not irregular in the product drafting process. Product layouts can be viewed from which products are most commonly purchased at the same time, by implementing those rules[10].

As the research has been done [12] that uses a priori algorithm as a decision making process in determining the layout of the goods in Buaran Market Puwokerto. From the explanation above, researchers will conduct shopping cart analysis by utilizing data mining techniques using a method of association and apriori algorithm to determine the layout of goods using the transaction data Pandak Baturraden convenience store that can be used as a product sales strategy.

2. Method

The concepts of this research are as follows:

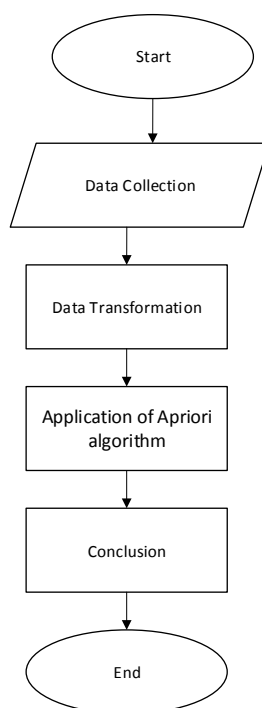


Fig. 1. Research Flowchart

In this study, researchers used the following research concepts:

1. Data Collection

In this research, researchers use primary data in the form of sales transaction data at Minimarket Pandak Baturraden. The Data consists of 829 transactions and 829 product names from 16 March to 17 April 2018. Each

transaction consists of 1 to 5 different product names. Preliminary data obtained by researchers in the form of a random sales receipt that has not been processed into the ready-made data in the research.

2. Data transformation

Then at this stage is the transformation of data to select data in order to get clean data and ready to use. Through the transformation process, researchers use one attribute, i.e. the name of the product shortened to the number 1 to 829. This data starts by converting data into a. Arff form by using the product code as an attribute and a "T" (true) that has a meaning purchased and a "?" sign that has a meaning of not being purchased.

3. Application of Apriori algorithm

In this stage, there are several forms of itemset ranging from 1-itemset, 2-itemset, up to K-Itemset. The process of forming itemset is done if multiple item sets have qualified minimum support. The support value is a value that shows the combination of items in the database taken from the number of transactions containing A divided by the number of transactions.

After the establishment of itemset that has qualified minimum support which will then be formed into a priori rule. Then after all the process of high frequency pattern is obtained next sought association rules that meet the minimum requirements for confidence by counting confidence associative rules $A \rightarrow B$. Confidence value is a value that indicates strong The relationship between items in the association rules that are addressed from the number of transactions containing A and B divided by the number of transactions containing A.

4. Conclusion

At this stage, researchers concluded that from the results of the research method the association has given a solution in decision making to determine the layout of the product in Pandak Baturraden minimarket by applying the combination of items that have been Generated.

3. Results and Discussion

3.1. Data Collection

This research uses primary data that is directly obtained from the Minimarket Pandak Baturraden. The data is a sales transaction data that must be input when the transaction process occurs. The transaction data is then saved for a transaction report. The Data used in the study began from 16 March to 17 April 2018 with total transaction and Total product name of 829 transactions and product name.

Table 4.1 Sample Data chunk Sales Transaction

Transaction	Purchased items
T1	SGM Explore 1 + Vanilla 150 GRM
	Djarum Super MLD Black 12
T2	Coca Cola 390 ml
	Max 32 GRM Driver
	BBQ Piattos 50 g
	Coffee Food All
T3	Prenagen Mommy Strawberry 200 g
	Magnum Mild 16
T4	Aqua 330 ml
	Smax Srz Cheese 40 g
	Sari Roti Cheese Sandwich
	Nissin Wafers Cheese 50 g
	Bread NN Rp. 5,000
T5	Cornetto Mini Disc
	Paseo Smart 250 Sht
	Frisian Flag Strawberry 115 ml
	Frisian Flag Brown 115 ml

3.2. Data Transformation

At this research, use Preprocessing/Cleaning data stages to discard data duplication, check data and correct errors on the data. Then give initials a product to facilitate the research process. The initializer is based on the brand or product name on each transaction. Then the brand/name of the product is adjusted based on alphabetical order to facilitate the time of the initializer process.

Table 4.2 Example of a Product Initial Cut

Initials Products	Product Name
1	3 in 1 Neopolitana 350 ml
2	ABC AA Blue Contents 2
3	ABC EXO Milk C 230 ml
4	ABC sweet soy sauce 135 ml
5	ABC Coffee Milk 31 g

After initializing all brand/product name, then enter the initials of each product in each transaction done on the sales transaction data.

4.3 Example of Transaction Data Cut

Transaction	Purchased items
T1	648.171
T2	137, 39,578,381
T3	602.437
T4	22,702,664,520,637
T5	141,569,222,289

Table 4.3 is the sales transaction data that has been given initials on each product purchased to make it easier to know how many items are purchased on each transaction. The next process is a tabulation of data that is the process of creating tables with members sign on each attribute of the product name is "T" (true) which means the product has been purchased and the mark "?" which means the product is not purchased by the consumer.

Table 4.4 Data Tabulation Format Sales Transaction

Transaction	Attribute												
	684	171	137	39	578	381	602	437	22	704	664	520	637
T1	Q	Q	?	?	?	?	?	?	?	?	?	?	?
T2	?	?	Q	Q	Q	Q	?	?	?	?	?	?	?
T3	?	?	?	?	?	?	Q	Q	?	?	?	?	?
T4	?	?	?	?	?	?	?	?	Q	Q	Q	Q	Q

Data in Table 4.4 then changed the data of the .arff form, .arff is *weka* format.

Relation: jajalan													
No.	1: 684 Nominal	2: 171 Nominal	3: 137 Nominal	4: 39 Nominal	5: 578 Nominal	6: 381 Nominal	7: 602 Nominal	8: 437 Nominal	9: 22 Nominal	10: 704 Nominal	11: 664 Nominal	12: 520 Nominal	13: 637 Nominal
1	t	t											
2			t	t	t	t							
3						t	t						
4								t	t	t	t	t	t

Fig. 2. Images form. Arff on Weka

Figure 2 is the . Arff format ready to be processed to start the process on Weka. In that format, there is a "T" mark indicating that the item has been purchased. In that format, there is also a product name attribute that has been replaced with the initial type of nominal product.

3.3. Application of Apriori algorithm

In this stage, the process is done from a priori algorithm to acquire a combination pattern of each item. In the process of forming an item using Weka must be determined in advance the support value and confidence value. This study used a support value of 0.35% and confidence by 60%. Support 0.35% stated that of all analyzed transactions showing the products purchased simultaneously and confidence 60% showed that if the buyer buys product 1 then there is a 60% chance of buying a product 2. With support and confidence value if inserted into Weka will produce the following data:

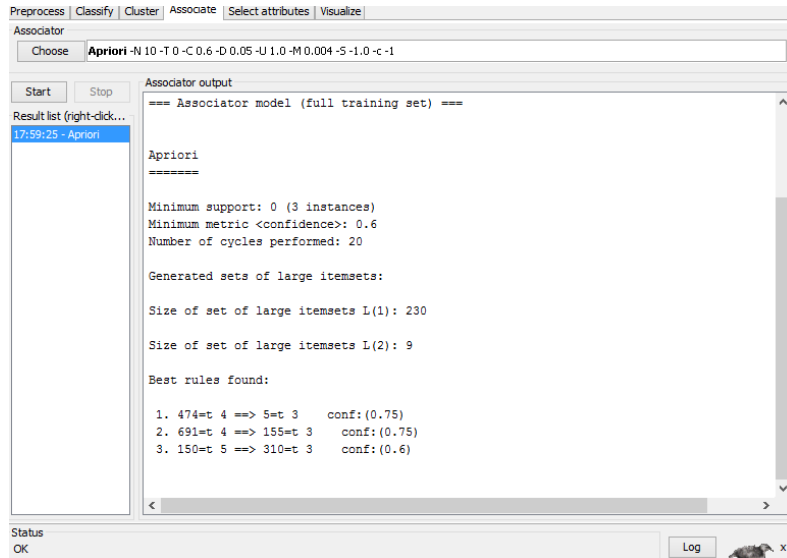


Fig. 3. Process results on Weka

Based on Figure 4.3, analysis generated using apriori algorithm on Weka with a minimum of 0.35% support and minimum confidence 60% There are 3 (three) combinations generated that have a relation to the sales transaction data is 474 □ 5 with confidence 75%, 691 □ 155 with confidence 75%, and 150 □ 310 with confidence of 60%.

a. Analysis High-frequency pattern

Furthermore, the process of forming 1-itemset combination with minimum support amounted to 0.35% and total transaction of 829 transactions. The process of specifying a support value can be done using the formula as follows:

$$Support (A) = \frac{Total\ Transaction\ Including\ A}{Total\ Transaction}$$

Table 5. Example 1-item combination snippet

Product Code	Product Name	Amount	Support (A)
2	ABC AA Blue Contents 2	5	5/829 = 0.60%
5	ABC Coffee Milk 31 g	7	7/829 = 0.84%
12	Adem Sari cans 320 g	4	4/829 = 0.48%
13	Ades 1500 ml	3	3/829 = 0.36%
14	Ades 350 ml	6	6/829 = 0.72%
15	Ades 600 ml	7	7/829 = 0.84%
21	Aqua 1500 ml	3	3/829 = 0.36%
22	Aqua 330 ml	8	8/829 = 0.97%
23	Aqua 600 ml	23	23/829 = 2.77%
24	Aqua 750 ml	6	6/829 = 0.72%

After the establishment of a 1-item combination in the 4.5 table with minimum support of 0.35% can be obtained as many as 230 items that meet the minimum support requirement. The next step is to form a 2-item by combining one item with another item.

The 2-item forming process can be obtained by using the formula as follows with a total transaction of 829 transactions:

$$Support(A, B) = P(A \cap B) = \frac{Sum\ of\ Transaction\ including\ A\ and\ B}{Total\ Transaction}$$

Table 5. 2-Item combination

Product Code	Product Name	Amount	Support
5 → 474	→ Milo Active 18 gm Milk Coffee	3	3/829 = 0.36%
150 → 310	Dancow 1 + Honey 200 gm → Ice Cream corneto	3	3/829 = 0.36%
155 → 691	Davos Strong 10 gm → siip Roasted 6.5 gm	3	3/829 = 0.36%
280 → 648	Gudang Garam Filter 12 → Sampoerna Mild 16	3	3/829 = 0.36%
282 → 381	Warehouse Garam Surya → Coffee All	3	3/829 = 0.36%
282 → 430	The warehouse of → Surya Luwak White koffie 20 gm	3	3/829 = 0.36%
292 → 560	Happy fruti → Paddle Pop Trico	3	3/829 = 0.36%
347 → 404	Fire ship steady 25 g → LA Lights 16	3	3/829 = 0.36%
350 → 430	Special Fire ship Mix 25 g → Mongoose White koffie 20 gm	4	4/829 = 0.48%

Table 6 is the result of a 2-item combination that satisfies 0.35% support for 9 combinations of items. The next step is to form a 3-item by combining one item with another item. The following is a 3-item combination table.

Table 7. 3-Item combinations

Product Code	Product Name	Amount	Support
282 → 350 → 430	Warehouse Solar → Salt ship Special Fire vessel 25 g → Mongoose White koffie 20 gm	2	2/829 = 0.24%
350 → 430 → 648	Special Mix Fire Boat 25 g → Luwak White koffie 20 gm → Sampoerna Mild 16	1	1/829 = 0.12%

Because no itemset meets the support value, then the formation of combinations is discontinued. Next count confidence values on each 2-item combination.

b. Establishment of association rules

Association rules are formed based on the selected itemset. The establishment of an association rule is done by calculating the minimum confidence value and looking for a combination of items that meet the minimum confidence of 60%. The confidence value can be searched by using the formula as follows:

Table 8. Association Rules of 2-itemset

A → B	Product Name	Support	Confidence
5 → 474	ABC Coffee Milk 31 g → Milo Active 18 gm	3/829 = 0.36%	3/7 = 42.86%
474 → 5	Milo Active 18 gm → ABC coffee milk 31 g	3/829 = 0.36%	3/4 = 75%
150 → 310	Dancow 1 + Honey 200 gm → Ice Cream corneto	3/829 = 0.36%	3/5 = 60%
310 → 150	Ice Cream corneto → dancow 1 + Honey 200 gm	3/829 = 0.36%	3/8 = 37.5%
155 → 691	Davos Strong 10 gm → siip Roasted 6.5 gm	3/829 = 0.36%	3/6 = 50%
691 → 155	SIIP Roasted 6.5 gm → Davos Strong 10 gm	3/829 = 0.36%	3/4 = 75%
280 → 648	Gudang Garam Filter 12 → Sampoerna Mild 16	3/829 = 0.36%	3/8 = 37.5%
648 → 280	Sampoerna Mild 16 → Warehouse Salt Filter 12	3/829 = 0.36%	3/24 = 12.5%
282 → 381	Warehouse Garam Surya → Coffee All	3/829 = 0.36%	3/14 = 21.43%
381 → 282	Coffee food All → Warehouse Solar salts	3/829 = 0.36%	3/22 = 13.64%
282 → 430	The warehouse of → Surya Luwak White koffie 20 gm	3/829 = 0.36%	3/14 = 21.43%

430 → 282	Luwak White koffie 20 grm → Warehouse Solar Salt	3/829 = 0.36%	3/7 = 42.86%
292 → 560	Happy fruti → Paddle Pop Trico	3/829 = 0.36%	3/8 = 37.5%
560 → 292	Paddle Pop trico → Happy fruti	3/829 = 0.36%	3/12 = 25%
347 → 404	Fire ship steady 25 g → LA Lights 16	3/829 = 0.36%	3/9 = 33.33%
404 → 347	LA Lights 16 → Fire ship steady 25 g	3/829 = 0.36%	3/13 = 23.08%
350 → 430	Special Fire ship Mix 25 g → Mongoose White koffie 20 grm	4/829 = 0.48%	4/20 = 20%
430 → 350	Luwak White koffie 20 grm → fire boat Special Mix 25 g	4/829 = 0.48%	4/7 = 57.14%

In table 4.8, Colored lines are indicating that there are three final association rules formed:

- a) 474 → 5, i.e. "Milo Active 18 grams" □ "ABC coffee milk 31 g" with a support value of 0.36% and confidence 75%. Meaning that buyers who bought Milo Active 18 grams have the possibility of 75% to buy ABC coffee milk of 31 g and Milo Active items 18 grams, and ABC coffee milk of 31 g is likely bought simultaneously at 0.36% of the existing 829 transactions.
- b) b) 150 → 310 i.e "Dancow 1 + Honey 200 grm" □ "Ice Cream Corneto" with a support value of 0.36% and confidence 60%. Meaning the buyer who bought Dancow 1 + Honey 200 GRM has the possibility 60% to buy Ice Cream Corneto and Items Dancow 1 + Honey 200 grm and Ice Cream Corneto is likely bought simultaneously at 0.36%.
- c) 691 → 155 is "SIIP Roasted 6.5 grams," □ "Davos Strong 10 GRM" with a support value of 0.36% and confidence 75%. Meaning the buyer who bought SIIP Roasted 6.5 grams has a probability of 75% to buy Davos Strong 10 grm and SIIP Roasted 6.5 items and Davos Strong 10 GRM is likely bought simultaneously at 0.36%.

3.4. Conclusion

Based on the results of the database, scanning can be generated 3 best final rules consisting of 829 transaction data and 829 attributes with minimum support of 0.35% and minimum confidence 60%. Items with product name Milo Active 18 grams have a history purchased as much as 4 times the purchase and 3 times purchased simultaneously with the product ABC coffee milk 31 g with confidence 75%.

Then the item with the product name Dancow 1 + Honey 200 GRM has a history purchased as much as 5 times the purchase and 3 times purchased simultaneously with the product Ice Cream Corneto with confidence 60%. Moreover, the item with the product name SIIP Roasted 6.5 grams has a history purchased as much as 4 times and 3 times purchased simultaneously with the product Davos Strong 10 grm with confidence 75%.

This shows that buyers have bought the 6 items of the product at the Pandak Baturraden minimarket. Utilizing the formed association rules can be used to determine the layout of the adjacent goods to facilitate the buyer in search of products related to each other and can be used as a reference in Stock of available goods.

4. Conclusion

Based on the research that has been done, can be concluded as follows:

1. There are 3 rules of the highest final association formed that meets the value of support and confidence value is:
 - a. If buying Milo Active 18 GRM, then it will buy ABC coffee milk 31 g with support 0.36% and confidence 75%.
 - b. If buying Dancow 1 + Honey 200 GRM, then it will buy Ice Cream Corneto with support 0.36% and confidence 60%.
 - c. If buying SIIP Roasted 6.5 grm, then it will buy Davos Strong 10 grm with support 0.36% and confidence 75%.

2. With the rules formed above, then it is advisable to determine the layout of goods, i.e. Milo Active 18 grm adjacent to the ABC coffee milk 31 G, Dancow 1 + Honey 200 grm adjacent to Ice Cream Corneto, and Siip Roasted 6.5 grm adjacent to Davos Strong 10 grm.
3. From the results of the study, Pandak Baturraden can use the association rules that formed to establish a sales strategy that is to determine the layout of the goods closely because it has a relation to one another.

References

- [1] L.Wang et al., "Efficient Mining of Frequent Item Sets on Large Uncertain Databases," IEEE Transactions on Knowledge and Data Engineering, vol.24, no. 12, pp. 2170 – 2183, Dec. 2012.
- [2] V.S.Tseng, "Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases," IEEE Transactions on Knowledge and Data Engineering, vol.25 , no. 8, pp. 1772-1786, Aug. 2013.
- [3] Tassa T., "Secure Mining of Association Rules in Horizontally Distributed Databases," IEEE Transactions on Knowledge and Data Engineering, vol. 26 , no. 4, pp.970-983, April 2014.
- [4] W.cao, "Research of The Mining Algorithm Based on Distributed Database," IEEE 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp.1252-1256, 2011.
- [5] F. Giannotti et al., "Privacy-Preserving Mining of Association Rules from Outsourced Transaction Databases," IEEE Systems Journal, vol.7, no.3, pp.385-395, SEPTEMBER 2013.
- [6] Z. Zhao, D. Yan, and W. Ng, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, pp. 1171-1181, May 2014.
- [7] S.Barua and J.Sander, "Mining Statistically Significant Co-location and Segregation Patterns," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, pp. 1185-1199, May 2014.
- [8] N. Liet al., "Implementation of Apriori Algorithm Based on MapReduce," 13th IEEE- ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp. 236-241, 2012.
- [9] S. Bhatnagar, "Algorithm for Finding Association Rules in Distributed Databases," 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, pp. 915-920, 2012.
- [10] Z.Zong and Z. ping, "A parallel algorithm of frequent itemsets mining based on bit matrix," IEEE International Conference on Industrial Control and Electronics Engineering, pp. 1210-1213, 2012.
- [11] K. Shah and S.Mahajan, "Maximizing the Efficiency of Parallel Apriori Algorithm," IEEE International Conference on Advances in Recent Technologies in Communication and Computing, pp.107-109, 2009.
- [12] S.Einakian and M. Ghanbari, "Parallel Implementation of Association Rule in Data Mining," Proceedings of the 38th Southeastern Symposium on System Theory Tennessee Technological University Cookeville, TN, USA, March 2006, pp. 21-26.