
Application of the Apriori Algorithm to Purchase Patterns

Ferry Putrawansyah

feypuawansyah@gmail.com

Institut Teknologi Pagar Alam

Article Information

Submitted : 25 Oct 2022

Reviewed: 10 Feb 2023

Accepted : 16 Mar 2023

Keywords

Algorithm

Apriori

Increase Income

Store

Abstract

The purpose of this study is to produce a pattern of buying goods using the Apriori Algorithm at the Viona store. The problem faced by the Viona store is that the Viona store has not been able to predict the pattern of purchases of goods that are frequently purchased by consumers and has not been able to optimally analyze the products that are frequently purchased and the combination of purchases made by consumers so that the pattern of sales of goods cannot be predicted properly. One way is to do past data sheet analysis to find purchasing patterns so that later the products sold can be predicted properly. The apriori algorithm is very effective in finding relationship patterns of one or more item sets in large data sets so it is effective in calculating sales transaction data and finding patterns of combinations of consumer habits and being able to quickly make product sales patterns. The results of the process of applying the Apriori algorithm to sales data at the Viona Store through the RapidMiner application are the same as those applied to the system built and using sales transaction data in 2022 using a minimum support of 30% and a minimum of 30% trust. So from this study, information was obtained that the goods that were often purchased together during the month of May were matches and cigarettes with 100% confidence. And for the month of June, the Viona store can recommend packages in the store by looking at the results of the combination of these 3 items so that consumer buying patterns can be anticipated by providing stock of goods that are frequently purchased by consumers..

A. Introduction

The development of information technology has significantly changed the conventional economic system into a digital economic system. This digital system allows the business world to conduct transactions using electronic media that offer more convenience, speed, and efficiency. Currently, there are many modern trade transactions that use technology as a medium or what is commonly called electronic transactions or electronic commerce or electronic commerce (With the existence of trading transactions every day, the data will increase over time. This encourages experts to further develop technology to help human work or even exceed human work. In this day and age, there are a lot of technologies that are able to adopt the processes and ways of thinking of humans, one of which is Data Mining. In recent years data mining has attracted the attention of the world of information systems, because of the availability of large amounts of data. Data mining is the process of extracting information or something important or interesting from the data in the database so as to produce very valuable information. The purpose of data mining is to find relationship patterns based on a complex set of data. Data mining consists of several methods, one of which is the Association method. The association method or association rule mining is a data mining technique to find associative rules between a combination of items. The association technique is useful in determining the highest frequency pattern between itemset sets. To find high-frequency patterns, the algorithm that is often used is the Apriori algorithm.

The apriori algorithm is one of the classical data mining algorithms. The a priori algorithm is a data retrieval algorithm with associative rules to determine the associative relationship of a combination of items (Damanik, 2021). Apriori algorithm is an algorithm used to find a pattern of relationships between one or more items in a dataset. The a priori algorithm is widely used in sales transaction data or commonly known as the market basket. Market basket itself is a technique of data mining that studies consumer behavior in buying goods simultaneously (Tri Susilo et al., 2019). The market basket is an itemset that is purchased simultaneously by customers in a transaction. So the a priori algorithm is often used in sales transactions, such as selling groceries

Based on the results of observations and interviews conducted by researchers at the Viona Store, the current problem is that all processes in this store are still using a manual system, there is no computerized system, plus there is no strategy that can increase sales turnover at the Viona Store. So Viona Stores often fail in the face of business competition in the field of selling basic necessities. This should be overcome by looking at the data on goods that consumers often buy, because from the data on consumer purchases, they can determine what combinations of goods are often purchased and what goods are rarely purchased by consumers. From the problems that have been described, it is hoped that the application of the a priori algorithm can be used as a strategy to increase sales turnover at the Viona Store. This algorithm can help in combining the ingredients in the Viona Store based on sales data. This system will process the sales data using the Apriori Algorithm method and this method was chosen because it can combine frequently purchased ingredients.

So that the application of the a priori algorithm is a method to find the pattern of relationships between one or more items in a dataset. A priori algorithms are widely used in transaction data or commonly called market baskets, for example a supermarket has a market basket, with the a priori algorithm, supermarket owners can find out the buying pattern of a consumer, if a consumer buys items A, B, there is a 50% chance he will buy item C, this pattern is very significant given the transaction data so far.

B. Research Method

2.1 Apriori Algorithm

Apriori algorithm is a classic algorithm in data mining. This algorithm is used to see the intensity of the occurrence of itemset or frequent itemset and the relevant association rules. This algorithm is also designed to work on databases that contain very many transactions, such as goods brought by customers in stores [1]. Apriori algorithm is a data retrieval algorithm with association rules to determine the associative relationship of a combination of items. The relationship between a priori algorithm and association is that it can find two or more attributes and two or more objects. The a priori algorithm is a type of association rule in data mining (Simbolon, 2019). Based on the above definition, it can be concluded that the a priori algorithm is a data mining technique to find association rules between a combination of items.

2.2 How the Apriori Algorithm Works

How the a priori algorithm works is as follows: (1) Prepare data in the form of a transaction format, (2) Form a combination of 1 itemset and calculate the support value. The combination of 1 itemset that is formed is called a candidate combination of 1 itemset, (3) Set the minimum support value. In determining the minimum support value, there is no formula or standard rule, it's just that what needs to be remembered is that the higher the minimum support value, the fewer combinations formed and the lower the minimum support value, the more combinations formed. The minimum support value that has been set from the beginning remains valid and is the same value for all itemset combinations, (4) After setting the minimum support value, then determine the combination of 1 itemset that meets the minimum support. This combination is called a 1 itemset combination that meets the minimum support, (5) Form a combination of 2 itemset obtained from a combination of 1 itemset that meets the minimum support and calculate the support value. The combination of 2 itemsets is called a candidate combination of 2 itemsets. The rules for forming combinations starting from combination 2 apply the combination formula ($AB=BA$) and the formula ($AB\neq BA$) does not apply (7) Determine the combination of 2 itemsets that meet the minimum support. The selected combination is called a 2 itemset combination that meets the minimum support, (8) Form a combination of 3 itemsets obtained from a combination of 2 itemsets that meet the minimum support and calculate the support value. The combination of 3 itemset that is formed is called the candidate combination of 3 itemset, (8) Determine the combination of 3 itemsets that meet the minimum support. The selected 3 itemset combination is called a 3 itemset combination that satisfies the minimum support, (9) And so on. The formation of the itemset combination can stop at any time, it can be combined 1 or 2 and so on. The rule to stop the formation of combinations is if there is no combination of

itemset values that meet the minimum support or there is a combination that meets the minimum support but the itemset combination is only 1 so it cannot be combined with other itemset combinations, (10) The last is to form association rules and look for confidence values.

The apriori algorithm is the basic algorithm proposed by Agrawal and Srikant in 1994 to determine frequent itemsets for Boolean association rules. The A-priori algorithm belongs to the type Association Rules on data mining the rules that state associations between multiple attributes often called affinity analysis or market basket analysis. Association analysis or association rule mining is a technique data mining to find rules a combination of items as well as useful to find a relationship interesting hidden in the data set big ones [2]. One of interesting association analysis stage the attention of many researchers for produce efficient algorithms is a high-frequency pattern analysis (frequent pattern mining). The importance of an association can be known with two benchmarks, namely: support and confidence. Support (value support) is a percentage the combination of these items in database, while confidence (value certainty) is the strength of the relationship between-item in association rules. High frequency pattern at the stage where the formation of a combination of items that have been meet the minimum support requirements previously determined. The support value on one itemset can be obtained by using a formula below:

$$\text{Support (A)} = \frac{\text{The transaction amount contains a}}{\text{Transaction value}}$$

The formula above has meaning to determine the support value at one itemset the number of transactions that contains item A divided by the number of transactions in databases. 2-item support value formula:

$$\text{Support(A, B)} = \text{Support(A} \cap \text{B)} = \frac{\text{The transaction amount contains b}}{\text{Transaction value}}$$

In the above formula to determine support value on two itemsets or more, the number of transactions that contains items A and B divided by the total transactions that occurred on dataset. Steps to take after determining the support value on high frequency itemset ago association rules are formed express the strength of the relationship itemset combination in the transaction. To define association rules which is formed at least the itemset must has two candidates A and B. On the rules that are formed apply the law associative A -> B does not hold B -> A. To define the rule A->B used formula:

$$\text{Confidence} = \text{Support(A} \cup \text{B)} = \frac{\text{The transaction amount contains b}}{\text{The transaction amount contains a}}$$

An example of applying the a priori algorithm is illustrated in the following figure:

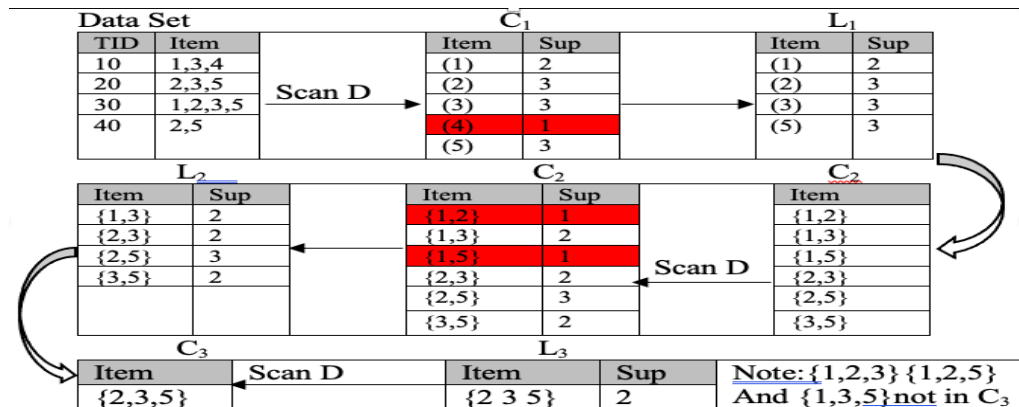


Figure 1: a priori algorithm

Based on the picture above, the minimum support used is 2. In the first iteration, items with support below 2 are eliminated from the L1-itemset. Then the candidate 2-itemset C₂ from the second iteration is formed from the cross product items in L1. After the 2 itemset candidates are calculated from the data, then 2 L₂ itemsets are determined. The same process is repeated in the 3rd iteration, but note that besides {2,3,5} which is a candidate 3-itemset C₃ there are actually itemsets {1,2,3} and {1,3,5} which can be obtained from the combination items in L₂, but the two itemsets are trimmed because {2,3} and {1,5} are not in L₂.

C. Result and Discussion

In making this a priori algorithm application system, the researcher first analyzes the sales data at the Viona Store using the RapidMiner application, with the aim of getting itemsets that are often purchased together which will later be used as product recommendations at the Viona Store so that it is expected to increase sales turnover at the Viona Store. The examples of data used are as follows:

Table 1. Data Association

Id_Transaction	Item of Buying
T1	Rice, Sweet Soy Sauce, Egg
T2	Cigarettes, Mineral water
T3	Rice, Cooking Oil, Egg, Garlic
T4	Cooking Oil, Sugar, Wafers
T5	Sugar, Ground Coffee, Cigarettes, Cooking Oil

From the data above, data transformation will be carried out based on the goods purchased. The results of the data transformation that has been carried out are as follows:

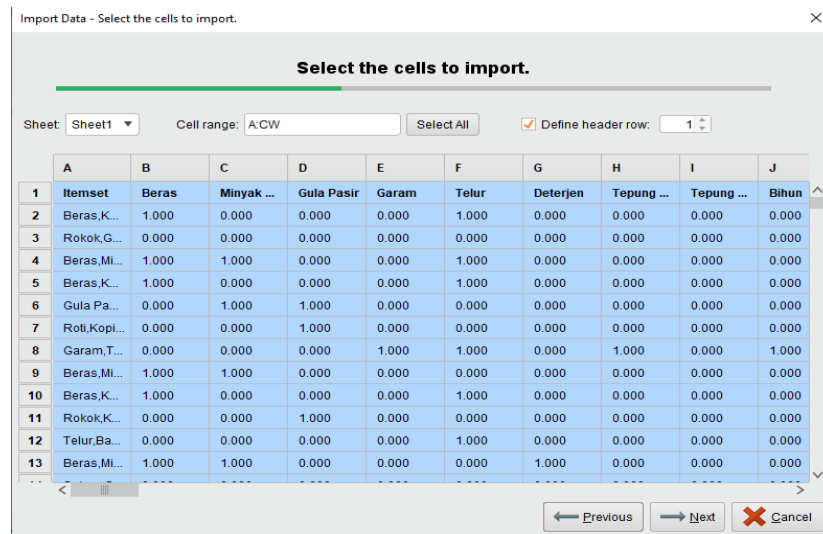


Figure 2. Data transformation

After preparing the dataset, the next step is to determine the calculation design on RapidMiner

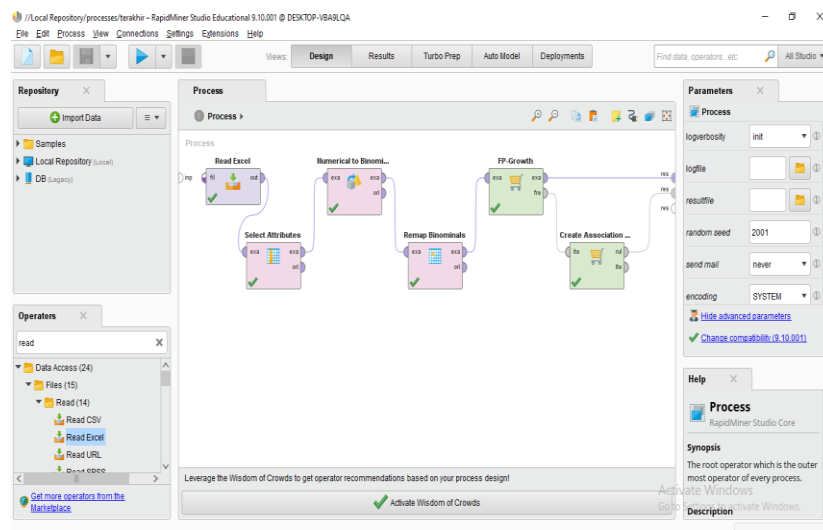


Figure 3. Determine the calculation design

After making the design and there are no errors, the design will run right away. After that, the results of the calculations can be seen starting from the data, graph and description. after that we determine the minimum value of support and confidence looks like the picture below

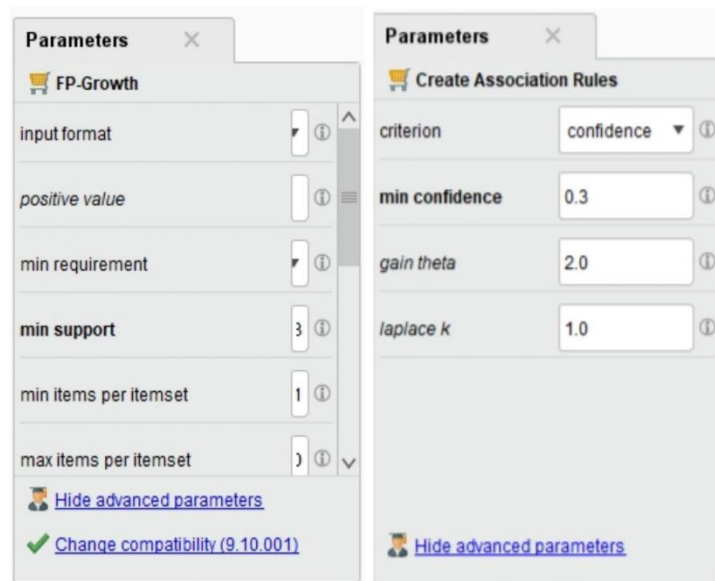


Figure 4. The minimum value of support and confidence

After that we determine, then we run it on rapidminer, we will get the conclusion of the itemset pattern in the image below

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	LIFT	ConvicL...
1	Minyak Goreng	Tepung Gandum	0.069	0.315	0.877	-0.370	0.043	2.854	1.287
2	Telur	Kecap Manis	0.079	0.412	0.905	-0.305	0.047	2.479	1.419
3	Telur	Mie Instan	0.081	0.423	0.907	-0.303	0.057	3.335	1.513
4	Kecap Manis	Telur	0.079	0.476	0.925	-0.253	0.047	2.479	1.542
5	Kecap Manis	Saos	0.083	0.500	0.929	-0.250	0.064	4.430	1.774
6	Minyak Goreng	Beras	0.113	0.514	0.912	-0.327	0.077	3.124	1.718
7	Tepung Gand...	Minyak Goreng	0.069	0.583	0.956	-0.168	0.043	2.854	1.872
8	Mie Instan	Telur	0.081	0.641	0.960	-0.172	0.057	3.335	2.248
9	Beras	Minyak Goreng	0.113	0.687	0.956	-0.216	0.077	3.124	2.491
10	Rokok	Korek	0.063	0.696	0.975	-0.119	0.058	10.978	3.078
11	Saos	Kecap Manis	0.083	0.737	0.973	-0.143	0.064	4.430	3.168
12	Korek	Rokok	0.063	1	1	-0.063	0.058	10.978	*

Figure 5. The conclusion of the itemset pattern

After that, we look for conclusions based on the itemset data so that we get the following pattern:

Association Rules

- [Minyak Goreng] --> [Tepung Gandum] (confidence: 0.315)
- [Telur] --> [Kecap Manis] (confidence: 0.412)
- [Telur] --> [Mie Instan] (confidence: 0.423)
- [Kecap Manis] --> [Telur] (confidence: 0.476)
- [Kecap Manis] --> [Saos] (confidence: 0.500)
- [Minyak Goreng] --> [Beras] (confidence: 0.514)
- [Tepung Gandum] --> [Minyak Goreng] (confidence: 0.583)
- [Mie Instan] --> [Telur] (confidence: 0.641)
- [Beras] --> [Minyak Goreng] (confidence: 0.687)
- [Rokok] --> [Korek] (confidence: 0.696)
- [Saos] --> [Kecap Manis] (confidence: 0.737)
- [Korek] --> [Rokok] (confidence: 1.000)

Figure 6. Conclusions based on the itemset data

From the rules that have been generated from the sales sample data at the Viona Store for the period of May 2022 which consists of 505 transaction data, the results of the tests carried out are as follows (1) If the consumer buys Rice, it is likely that consumers will also buy Cooking Oil with Confidence 67.5% (2) If consumers buy Cooking Oil, it is likely that consumers will also buy Wheat Flour with Confidence 31.8%, (3) If consumers buy rice, it is likely that consumers will also buy wheat flour with 42.2 % Confidence (4) If consumers buy Wheat Flour, it is likely that consumers will also buy Rice with Confidence 58.3 % (5) If consumers buy eggs, it is likely that consumers will also buy Instant Noodles with Confidence 42.3 % (6) If consumers buy Instant Noodles, it is likely that consumers will also buy eggs with 64 % Confidence (7) If consumers buy eggs, it is likely that consumers will also buy sweet soy sauce with 41.2% Confidence (8) If consumers buy sweet soy sauce, it is likely that consumers will also buy eggs with 47.6% Confidence (9) If consumers buy sweet soy sauce, it is likely that consumers will also buy sauces with 50% Confidence (10) If consumers buy Sauce, it is likely that consumers will also buy Sweet Soy Sauce with Confidence 73.7 % (11) If the consumer buys cigarettes, it is likely that consumers will also buy lighters with confidence 69,6%(12) If the consumer buys lighters, it is likely that consumers will also buy cigarettes with 100% Confidence

D. Conclusion

The results of the process of applying the Apriori algorithm to sales data at the Viona Store via the RapidMiner application are the same as those applied to the system built and using sales transaction data for May 2022 using a minimum support of 30% and a minimum of 30% trust. So from this study, information was obtained that the goods that were often purchased together during the month of May were matches and cigarettes with 100% confidence. And for June, the Viona store can recommend packages in the store by looking at the results of the combination of these 3 items. Based on this, it can be seen that the purchasing patterns of consumers who shop at Viona stores so that later this pattern data can be used as information to prepare stocks of goods that are in accordance with the most consumer buying patterns.

E. Acknowledgment

The researcher would like to thank the Chancellor and P3M of the Pagar Alam Institute of Technology who have provided moral and material support.

F. References

- [1] Damanik, Yeusi, F. S. Y., Sumarno, Indra, G., Dedy, H., & Kirana Ika Okta. (2021). *Penerapan data mining untuk pengelompokan penyebaran covid-19 di sumatera utara menggunakan algoritma k-means*. 1(2), 109–132.
- [2] Informasi, S., & Triguna Dharma, S. (2019). *Penerapan Data Mining Dalam Menganalisa Data Penjualan Untuk Mendapatkan Pola Rekomendasi Penjualan Menggunakan Algoritma Apriori Pada K3 Mart * Dicky Nofriansyah, Milfa Yetri, Kamil Erwansyah, Suharsil*. 18(SAINTIKOM), 176–182.
- [3] Kurniawati, A. (2014). Pemetaan Pola Hubungan Program Studi Dengan Algoritma Apriori. *Edu Komputika Journal* , 51-58.

- [4] N. C. Dewi, F. Putrawansyah, and D. Puspita, "Implementasi Algoritma FP-Growth Pada E-Commerce Kopi Pagar Alam Menggunakan Framework Codeigniter," *Indones. J. Comput. Sci.*, vol. 10, no. 2, 2021.
- [5] N. C. D. Ferry Putrawansyah, "Implementasi Algoritma K-Means dengan menggunakan Metode Profile Matching Pada Alumni STT Pagar Alam," *Jurnal Ilmiah Binary*, vol. 3 No 2, 2021.
- [6] Rahmi, A. N., Mikola, Y. A., & Kunci, K. (2021). *Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus : Toko Bakoel Sembako)*. 4(1).
- [7] Simbolon, P. H. (2019). Implementasi Data Mining Pada Sistem Persediaan Barang Menggunakan Algoritma Apriori (Studi Kasus : Srikandi Cash Credit Elektronik dan Furniture). *Jurnal Riset Komputer*, 6(4), 401–406.
- [8] Tri Susilo, A. A., Sunardi, L., & Waruwu, Y. (2019). Penerapan Algoritma Apriori Pada Data Penjualan Kosmetik Di Toko Sharly Kota Lubuklinggau. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 4(2), 92–100. <https://doi.org/10.32767/jutim.v4i2.648>
- [9] E. Irfiani, "pplication of Apriori Algorithms to Determine Associations in Outdoor Sports Equipment Stores," *Journal Publications & Informatics Engineering Research Sinkron*, Vols. Volume 3, Number 2, April 2019, pp. 219 - 222, 2019.