



## Implementation of The Apriori Algorithm for Spare Parts Data Sales

Siti Nur Khasanah<sup>1</sup>, Levi Juniarti<sup>2</sup>, Sandra Jamu Kuryanti<sup>3</sup>, Eko Yulianto<sup>4</sup>

<sup>1234</sup>Information Systems, STMIK Nusa Mandiri, Jl. Kramat Raya No.18, RT.5/RW.7, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450

E-mail: [siti.skx@nusamandiri.ac.id](mailto:siti.skx@nusamandiri.ac.id), [levhy20@gmail.com](mailto:levhy20@gmail.com), [sandra.sjk@bsi.ac.id](mailto:sandra.sjk@bsi.ac.id), [eko.eui@bsi.ac.id](mailto:eko.eui@bsi.ac.id)

### ARTICLE INFO

### ABSTRACT

#### Article history:

Received: 12/07/2020

Revised: 22/08/2020

Accepted: 30/09/2020

#### Keywords:

Data Mining, Apriori Algorithm, Association rule, Confidence Value, Support value

*Pakubumi piling spare parts are very much needed when there is a lot of infrastructure development in the capital city and in small cities. Spare parts that are used at the time of construction and for stocks carried out by project workers that will be prepared by suppliers as well as in CV. Abadi Perkasa. With a lot of available sales data, an algorithm is needed to manage the sales data. Data processing sales data mining with a priori algorithm technique can look for the highest frequency pattern where the highest frequency value is above a certain threshold which is called minimum support. After getting the highest support value, then the confidence value is sought to determine the association rules for one product with another. Looking for a minimum of support and confidence can also be found with the help of rapidminer tools. From this research, it can be seen that the a priori algorithm can find out which products are often purchased as well as determine the patterns of relationships between products and help to develop future marketing strategies.*

Copyright © 2020 Jurnal Mantik.  
All rights reserved.

## 1. Introduction

In a company, supermarket, or even a small shop, it must have data stored in its database, so that the more data that is inputted the more pile up. Some companies, supermarkets, or small shops still manage this data manually. In an era of increasingly sophisticated analysts analyzing the data using a computerized application. In the retail business, one way that can be done to determine market conditions (customers) is to observe sales transaction data. Sales transaction data is stored in a server database and then this data is processed, resulting in sales reports and income statements. However, the sales data can be processed further so that new information can be obtained. CV. Abadi Perkasa is a shop engaged in piling equipment spare parts or often called heavy equipment spare parts. CV. Abadi Perkasa must meet the needs of the spare parts that the project will use for stock and the needs of the ongoing project. With this, stores are required to make decisions correctly and quickly in determining sales strategies. To find out what sales strategy is made, the store management can take advantage of the sales data information in each transaction to be able to find out the buying patterns of customers who are needed or that are often purchased. To analyze these buying patterns, you can use Affinity Analysis or what is called Market Basket Analysis.

There are more and more buyers every day as well as sales product data if the sales data is not managed properly it will become an accumulated archive and cannot help the progress of the store itself. With that, a store must have a system to manage sales data that is often purchased. Data management can be done using a priori algorithm, where the a priori algorithm can determine the minimum sales support and can analyze product sales patterns and determine the minimum confidence to find out the rule association pattern of product sales, the store can make a reference in developing sales strategies for the future [1].

The way to find out what spare parts are often purchased simultaneously, can use the association rule, is a data mining technique to find association rules between a combination of items. This association search process uses the aid of a priori algorithm, which is an algorithm used to produce an association rule with an "if then" pattern which functions to form possible item combinations, then tested whether the combination meets the minimum support and confidence parameters which are the threshold values given by user [2]. With this a priori algorithm, it will produce a combination pattern of items and rules as important knowledge and information from sales transaction data.

From the description of the background of the problem above, the writer makes the title "Implementation of the Apriori Algorithm for Data Analysis of Sales of Spare Part.



## 2. Literature Review

In previous research, a pharmacy has implemented an information system in processing incoming drug data, outgoing drugs and reporting. However, this pharmacy has not applied science to sales and purchase data as a reference for determining drug buying trends. So that the process creates obstacles for pharmacists in finding drugs that are often needed by consumers. Therefore, an a priori algorithm for consumer purchasing patterns is made to look for patterns of drug buying tendencies, so that information will be obtained about which drugs are often purchased and which drugs consumers buy simultaneously [3].

Subsequent research is a company that makes questionnaires to employees and employees of what game applications are most in demand. The a priori algorithm really helps determine what games are most interested or played, the data obtained from the questionnaire is processed, starting from the formation of support and confidence values, then it will produce a final association that meets the support and confidence values [4].

In the next research, a company analyzes what defects often appear in the injection process and the combination of defect items that often occur. The method to find this out is to use the a priori algorithm method so that it can manage input values in accordance with the criteria for items that have certain support and confidence values. A priori algorithm, one of the algorithms in data mining that is used in the association rule to determine an itemset which functions to help find patterns in data [5].

## 3. Research Method

The research methodology is the systematic stages of research to help the research be well directed. The following is the research methodology conducted by the author.

### a) Identify the Problem

The first step of this research is to identify the problems that exist in CV. Abadi Perkasa and carried out based on the formulation of the problem based on the background of the problem.

### b) Literature Study

The author studies and understands the theories used in this study. The theory is obtained through books, research journals, previous theses, and e-books.

### c) Data Collection

The author collects data using primary data and secondary data by means of the author coming directly to CV. Abadi Perkasa, observation by observing the sales process and conducting interviews with shop owners regarding matters related to research.

### d) Research data

In this study, the data used by the author are primary data and secondary data

### e) Data analysis

In data analysis, the writer analyzes using a priori algorithm to find support value and confidence value.

### f) Research results

After the authors get the value of support and confidence values, they can confess which products are most interested in and which products are interrelated with one another.

### g) Conclusions and suggestions

This stage is the final stage the author concludes and provides suggestions on existing problems.

## 4. Discussion

### 4.1 Determining the Population and Sample Research

The population in this study were Pakubumi spare parts, amounting to 577 products. The sample is part of the number and characteristics of the population. The sampling technique used by the author in this study is simple random sampling technique (simple random sampling). To determine the number of samples the author uses the Slovin formula as follows:

$$n = \frac{N}{1 + N \cdot (\epsilon^2)}$$

Based on the Slovin formula above, obtain the following sample sizes:

$$n = \frac{N}{1 + N \cdot (e^{-2})} = \frac{577}{1 + 77 \cdot (0.1^2)}$$

$$= \frac{577}{1 + 577 \cdot (0.01)} = \frac{577}{1 + 5.77} = \frac{577}{6.77} = 85.2$$

By calculating using the Slovin formula above which has a critical value of 10%, the required sample size is 85.2 and rounded up to be 85 product data.

**4.2 Making Tabular Format**

Tabular format of sales data for each invoice number, if formed will look like Table 1 following:

**Table 1.**

Tabular Format Sales Data

	13T16F1	15Z215	17Z31 7	.....	30017- 90100	4045Z1	5Z3	5Z89	BLA 1007	BLA 1157	BLA 1160
1	1	0	1	.....	1	0	1	0	0	0	1
2	0	0	0	.....	0	0	0	0	0	0	0
3	1	1	1	.....	0	0	0	0	0	0	0
4	0	0	0	.....	0	0	0	0	0	1	0
5	1	0	1	.....	0	0	1	0	0	0	0
6	0	0	0	.....	0	0	1	0	0	0	0
7	0	1	1	.....	1	0	0	0	0	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
485	0	0	0	.....	1	0	0	1	0	0	0
486	0	0	0	.....	0	0	1	0	0	0	0
487	0	0	1	.....	0	0	0	0	0	0	0
488	0	1	0	.....	0	0	1	0	0	0	0
489	0	0	0	.....	0	0	0	0	0	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
502	0	1	0	.....	1	0	0	0	0	0	0
503	0	0	0	.....	0	0	1	0	0	0	0
504	0	0	1	.....	0	0	0	1	0	0	0
505	0	0	1	.....	1	0	1	0	0	0	0
506	0	1	0	.....	0	0	0	0	0	0	0

Tabular tables provide descriptions of the values in the table. The table above has columns (fields) and rows (records). Tabular tables are also used to make it easier for researchers to find out the number of transactions in each product, where 1 indicates transactions or products sold while 0 is the absence of transactions on that product.

**4.3 Analyzing the Highest Frequency Pattern**

In this step the researcher took the minimum support value was 0.200 / 20%. The determination of the values is adjusted to produce the correct association rules.

**Table 2.**

Establishment of 1 Set Item

NO	PRODUK	JUMLAH	SUPPORT	SUPPORT%
1	17Z317	104	0.206	21%
2	18T6702	104	0.206	21%
3	18T6720	105	0.208	21%
4	20Z758/641	127	0.251	25%
5	219T150	101	0.200	20%
6	2425Z61	104	0.206	21%
7	2438U192F1 "NABCO	103	0.204	20%
8	2444Z1090F1	137	0.271	27%
9	30003-86100	115	0.228	23%
10	2438U43F1	178	0.352	35%
11	30017-90100	120	0.238	24%
12	4045Z10D120	100	0.198	20%
13	5Z3	130	0.257	26%
14	BLA 1007	115	0.228	23%
15	2445Z65	104	0.206	21%
16	OIX 1841	104	0.206	21%
17	20K81D9	100	0.198	20%
18	20K81D5	130	0.257	26%
19	DD25	115	0.228	23%



Table 2 is sales data with a minimum support of 0.200 (20%). Of the 85 sales data that meet the support value, only 19 product items, then this product data will be formed into 2 Itemset. The resulting minimum support candidates are 168, itemset products that meet Minimum support of 0.200 (20%) will be selected, those that do not meet the minimum support will be eliminated. As shown in table 3.

**Table 3.**

Minimum Support 2 ItemSet 0.200/20%			
PRODUK	JUMLAH	SUPPORT	SUPPORT(%)
17Z317 DAN 2445Z65	104	0.206	21%
18T6702 DAN OIX 1841	104	0.206	21%
5Z3 DAN 20K81D5	130	0.257	26%
BLA 1007 DAN DD25	115	0.228	23%

**4.4 Establishment of Association Rules**

After looking for the highest frequency pattern, then look for the association rule with a minimum confidence value. to calculate the confidence value with the rule A → B.

**Table 4.**

Candidate Of Associaton			
NO	PRODUK	JUMLAH	CONFIDENCE
1	17Z317 dan 2445Z65	104	1.000
2	18T6702 dan OIX 1841	104	1.000
3	5Z3 dan 20K81D5	130	1.000
4	BLA 1007 dan DD25	115	1.000

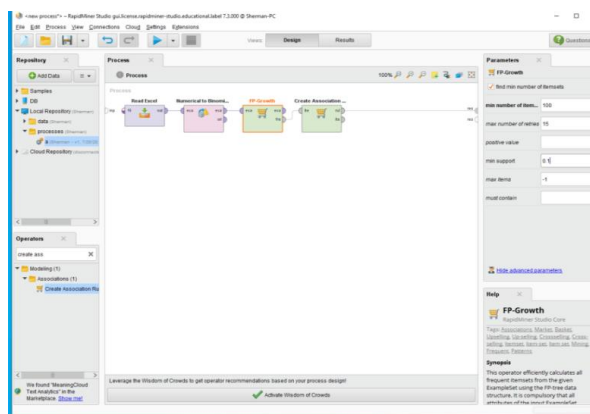
The 2-itemset in the table is taken from products that meet the minimum 2-itemset support or the highest frequency pattern. Based on the prospective association rules in the table, those that meet a minimum support of 0.200 and a minimum confidence of 1,000 can be seen in the table below:

**Table 5.**

Association Final		
ATURAN ASSOSIASI	SUPPORT	CONFIDENCE
JIKA MEMBELI 5Z3 MAKA AKAN MEMBELI 20K81D5	0.257	1
JIKA MEMBELI 20K81D5 MAKA AKAN MEMBELI 5Z3	0.257	1
JIKA MEMBELI DD25 MAKA AKAN MEMBELI BLA 1007	0.228	1
JIKA MEMBELI BLA 1007 MAKA AKAN MEMBELI DD25	0.228	1
JIKA MEMBELI OIX 1841 MAKA AKAN MEMBELI 18T6702	0.206	1
JIKA MEMBELI 18T6702 MAKA AKAN MEMBELI OIX 1841	0.206	1
JIKA MEMBELI 2445Z65 MAKA AKAN MEBELI 17Z317	0.206	1
JIKA MEMBELI 17Z317 MAKA AKAN MEMBELI 2445Z65	0.206	1

Based on the table above, it can be seen that the Association Rules are if you buy 5Z3 (bushings) you will buy 20K81D5 (rivet) and vice versa, if you buy DD25 (ring) you will buy BLA 1007 (oil seal) and vice versa, if you buy OIX 1841 ( oil seal) then you will buy 18T6702 (retainer) and vice versa, if you buy 2445Z65 (piston cup) you will buy 17Z317 (lining) and vice versa.

The following is the implementation of the Apriori Algorithm using Rapid Miner:



**Fig 1.** Apriori Algorithm implementation using Rapid Miner

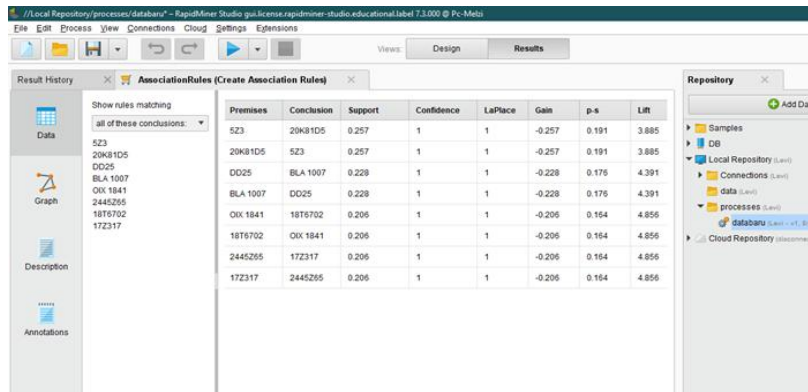


Fig 2. Display the results of Support and Confidence value data

Figure 2 shows the results of a minimum Support value of 0.200 and a confidence value of 1.0 from data processing using the rapidminer tools, products that meet the minimum support value are 5Z3 (bushing) dan 20K81D5 (rivet) (0.257/1.0), DD25 (ring) dan BLA 1007 (oil seal) (0.228/1.0), OIX 1841 (oil seal) dan 18T6702 (retainer) (0.206/1.0) 2445Z65 (piston cup) dan 17Z317 (lining) (0.206/1.0).

### 5. Conclusions

From the research results, it was concluded that the a priori algorithm was implemented in spare part sales, namely if you buy 5Z3 (bushing) you will buy 20K81D5 and vice versa, if you buy DD25 (ring) you will buy BLA 1007 (oil seal) and vice versa, if you buy OIX 1841 (oil seal) will buy 18T6702 (retainer) and vice versa, if you buy 2445Z65 (piston cup) you will buy 17Z317 (lining) and vice versa. With this implementation, it can be concluded that by knowing the pattern of customer purchases with a lot of buying items 5Z3, 20K81D5, DD25, BLA 1007, OIX 1841, 18T6702, 2445Z65, 17Z317. So the shop can increase the stock of these items and items that are less desirable to the shop's customers can minimize the inventory stock so that the stock doesn't accumulate in the warehouse.

While the suggestion proposed is for the next research it is suggested to increase the research time, not enough, 3 months should be added by 1 year or 2 years so that the research is more accurate. And it is hoped that researchers can develop or build an application that can be used for data processing, in the implementation of selling product items not only using a priori algorithm but can use other algorithms to be a comparison. To use the tools, not only using the rapidminer tools, you can also use other tools so that the output obtained is more accurate.

### 6. References

- [1] J. Speed and S. P. Engineering, "Journal Speed – Sentra Penelitian Engineering dan Edukasi – Volume 10 No 4 – November - 2018," vol. 10, no. 4, pp. 90–95, 2018.
- [2] D. Listriani, A. H. Setyaningrum, and F. E. M.A, "Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Pola Belanja Konsumen ( Studi Kasus Toko Buku Gramedia Bintaro )," *J. Tek. Inform. Vol 9 No. 2, Univ. Islam Negeri Jakarta*, vol. 9, no. 2, pp. 120–127, 2016.
- [3] A. Ishaq, L. A. Utami, and S. Mariana, "Analisa Pola Penjualan Obat Menggunakan Algoritma Apriori Pada Apotek Zam-Zam Bogor," vol. 08, no. 1, pp. 13–23, 2019.
- [4] A. Rachman and S. Wasiyanti, "Pengukuran Kualitas E-Commerce Shopee Terhadap Kepuasan Pengguna," *Paradig. J. Komput. dan Inform. Univ. Bina Sarana Inform.*, vol. 21, no. 2, pp. 143–148, 2019, doi: 10.31294/p.v20i2.
- [5] O. Pahlevi, A. Sugandi, and I. D. Sintawati, "Penerapan Algoritma Apriori Dalam Pengendalian Kualitas Produk," *Sinkron*, vol. 3, no. 1, pp. 272–278, 2018.