Rulinawaty

by Rulinawaty Rulinawaty

Submission date: 27-Aug-2021 02:06PM (UTC+0700)
Submission ID: 1636711567
File name: Implementation_of_Data_mining_with_Apriori_Techniques.pdf (657.44K)
Word count: 3562
Character count: 18474

Implementation of data mining with Apriori techniques to determine the pattern of purchasing of agricultural equipment (Case Study: XYZ Store)

Implementation of data mining with Apriori techniques to determine the pattern of purchasing of agricultural equipment (Case Study: XYZ Store)

Supriyono[1], Kiki Farida Ferine[2], Diana Puspitasari[3], Rulinawaty[4], and Elkana Timotius[5]

[1]Universitas Muria Kudus, Kudus, Indonesia. Email: supriyono.si@umk.ac.id
[2]Universitas Pembangunan Panca Budi, Medan, Indonesia. Email: kikifarida@dosen.pancabudi.ac.id
[3]Universitas Diponegoro, Semarang & Universitas Dian Nuswantoro, Semarang
[4]Universitas Terbuka, Indonesia. Email: ruly@ecampus.ut.ac.id
[5]Department of Industrial Engineering, Universitas Kristen Krida Wacana, Jakarta, Indonesia. Email: mr.elkana@gmail.com

Abstract. The aim of this research is to take advantage of data mining techniques using the Apriori sales data algorithm for agricultural equipment. This is due to the fact that the high potential agricultural sector in Indonesia has obstacles to the use of technology. Research was conducted in the district of Simalungun in a shop selling agricultural needs. Sales activities in the store continue and generate more and more data. In order to be of use to the resulting data, it is necessary to process these data with a certain algorithm that provides great benefits, in particular by maximizing the sales profits of agricultural products. Apriori Algorithm is one of the methods of data mining, the activities of which include data collection and the use of old data with the aim of finding regularities, patterns or relationships in data collection. The output of the algorithm can help future decision-makers, where one of the advantages is the rearrangement of the product layout, such as the most frequently sold products being assembled so that they are easily visible to consumers and can properly prepare stock items for the store.

Keywords: Analysis, Data Mining, Rule Association, Apriori, Agriculture

1. Introduction

Indonesia is known as a country with a high potential for agricultural yield. This is demonstrated by the agricultural sector, which is currently not only an area of employment, but also a food producer. The emergence of industry and small and medium-sized enterprises generated significant foreign exchange. However, the obstacle that arises, particularly in this technological era, is that the agricultural sector is faced with a problem of lack of use of technology and information. This affects the performance and profits of the agricultural sector [1][2]. Overcoming this problem, the Government is providing guidance to a number of small and medium-sized enterprises that have been set up to become a strong foundation for supporting the country's growing economy, in particular the agricultural sector in today's modern technology era by managing agribusiness and food processing using technology, information and communication. One shop engaged in business in the agricultural sector, which supplies different types of seeds and other agricultural needs, has problem recording every sale transaction where there is no data on the types of products available because there are so

many types of products sold. This research was carried out in the district of Simalungun in one of the shops, which is a place to shop for agricultural equipment and needs that are complete, high quality and cheaper than its competitors. There are also a lot of consumers in this store, which results in a lot of data piled up in the transaction process, so that there is often confusion between customers to find the place for the product they are looking for simultaneously. On the basis of the above problems, we need a system that is able to combine products that are often purchased at the same time. Data from each transaction carried out in cash can be analyzed in order to identify future sales strategies using artificial intelligence applications [3]–[5]. Data mining is one of them [6]–[8]. This is very useful for stores in the determination of product stock supplies, in the enforcement of discounts, and in helping to identify layouts that can make it easier to find items that are often purchased simultaneously.

Data mining is [9] the process of analyzing the data to find a pattern from the data set. Data mining is capable of analyzing large data[10], [11] into information in the form of patterns that are meaningful to decision support. One of the mining data techniques that can be used is association data mining [12], [13] or what is commonly referred to as market basket analysis. The market basket is defined as an item that is purchased simultaneously by a customer in a transaction [14]. Market basket analysis is a powerful tool for the implementation of a cross-selling strategy. This method begins with the search for a number of frequent items and continues with the formation of association rules. The Apriori [10] algorithm is one of the most popular algorithms for finding a number of frequent transaction data items stored in the database. This study uses the Apriori algorithm to help find a number of association rules in the sales transaction database of agricultural products in stores so that they can be used as consideration for the development of effective marketing and sales strategies.

2. Methodology

2.1. Association Rule Mining

The Association Mining Rule, which is one of the most important data mining techniques, is based on the identification of associations, correlations and patterns between items or items in the database. The aim of the Association Mining Rule is to discover strong rules of association between all frequent items. It can be made real through two procedures [15]. The first procedure involves the discovery of all frequent items, and the frequency of the item set should be at least equal to the minimum support threshold. The second procedure involves the establishment of a strong corresponding association rule based on the frequent item set obtained, a rule which must meet the minimum confidence threshold. Finding frequent patterns or set items in large databases is the core of any association rule discovery algorithm [16]–[18].

2.2. The Apriori Algorithm

Apriori tool has been applied effectively in most cases. market basket analysis, e-commerce recommender system and healthcare domain [19]. This method helps in identifying recurring characteristics in a large data set. This algorithm is very robust as it employs several rule-metrics. ‘Support', ‘Confidence' and ‘Lift' to differentiate intense associating rules with high correlation among them. Association rules and rule measurements will be discussed in the following subsections [15].

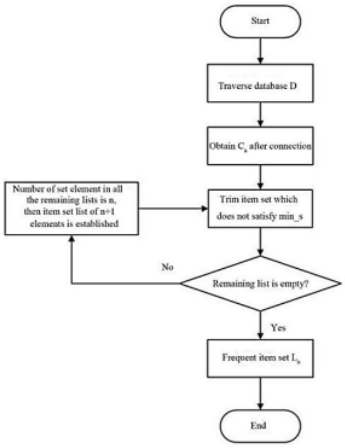

Figure 1. Flow diagram of the original Apriori algorithm.

3. Results and Discussion

Data needs to be processed when the a priori algorithm is used to determine the pattern of purchase of agricultural equipment. Sources of research data obtained from data collected through observational phases, interviews and literature studies in the district of Simalungun. In this study, an analysis of the sales data was carried out in stores with the aim of being able to see the most purchasing patterns so that they could supply the most purchased stock of goods so that the inventory was still there and to fix the combination of product arrangements to make it easier to find together as a strategic step towards satisfying the customer with the needs of the goods.

Table 1. Agricultural Product Sales Transactions

Transaction	Product name
1	Corn Seeds, Urea Fertilizer
2	Urea Fertilizer, Red Chili Seeds
3	Organic Fertilizer, Phonska Fertilizer, Green Chili Seeds
4	Gramoxon Pesticides, Kangkung Seeds, Green Mustard Seeds, Urea Fertilizer, NPK Fertilizer
5	SS Fertilizer, ZA Fertilizer
6	Cayenne Pepper Seeds, Push up Pesticides, Urea Fertilizer
7	Curly Chili Seeds, Red Chili Seeds
8	Urea Fertilizer, Gramoxon Pesticides, Corn Seeds
9	Basmilang Pesticides, ZA Fertilizers, ST
10	TSP Fertilizer, Urea Fertilizer, Pesticide Basmilang
11	Corn Seeds, Kale Seeds, Organic Fertilizer
12	Basmilang Pesticides, Gramoxon Pesticides, Tomato Seeds, Phonska Fertilizers
13	Urea Fertilizer, Phonska Fertilizer, Basmilang Pesticides, Corn Seeds
14	Gramoxon Pesticides, Phonska Fertilizers, Push up Pesticides
15	Corn Seeds, Tomato Seeds, Kale Seeds
16	Cayenne Pepper Seeds, ZA Fertilizer, SS Fertilizer
17	ST Fertilizer, Urea Fertilizer, Curly Chili Seeds
18	Phonska Fertilizer, Urea Fertilizer, Pesticide Basmilang
19	Mustard Green Seeds, Corn Seeds, Gramoxon Pesticides, Push up Pesticides
20	ZA Fertilizer, ST Fertilizer, Push up Pesticide, Urea Fertilizer, Phonska Fertilizer
21	Corn Seeds, Tomato Seeds, Organic Fertilizer
22	Corn Seed, Urea Fertilizer, ZA Fertilizer
23	Red Chili Seeds, Gramoxon Pesticides, SS Fertilizers, Basmilang Pesticides
24	Phonska Fertilizer, Urea Fertilizer, Pesticide Basmilang
25	Kale seeds, curly chili seeds, corn seeds

3.1. Analysis of High Frequency Patterns

At this stage, it begins by setting the minimum support with the aim of looking for the highest frequency pattern in each item combination, in which case the minimum support is 16%, then separating each item to look for the highest frequency pattern.

Table 2. List of Agricultural Product Sales Transactions

No	Item Name	Quantity
1	Corn Seed	9
2	Kale seeds	4
3	Green Mustard Seeds	2
4	Cayenne Pepper Seeds	3
5	Red Chili Seeds	3
6	Curly Chili Seeds	3
7	Tomato Seeds	3
8	Organic fertilizer	3
9	Urea fertilizer	12
10	Phonska fertilizer	7
11	NPK fertilizer	1
12	SS fertilizer	3
13	ZA fertilizer	5
14	TSP fertilizer	1
15	Fertilizer ST	3
16	Gramoxon pesticide	6
17	Basmilang Pesticides	7
18	Pesticide Push up	4

a) Based on table 2, look for the minimum value for iteration-1 Support using the following formula.

$$Support\ A = \frac{\sum Jumlah\ Transaksi\ Mengandung\ A}{\sum Total\ Transaksi} \quad (1)$$

The following is the support value data for each product as shown in Table 3.

Table 3. Item Set and Support Value

No	Item Name	Number of Items	Support
1	Corn Seed	9	36%
2	Kale seeds	4	16%
3	Green Mustard Seeds	2	8%
4	Cayenne Pepper Seeds	3	12%
5	Red Chili Seeds	3	12%
6	Curly Chili Seeds	3	12%
7	Tomato Seeds	3	12%
8	Organic fertilizer	3	12%
9	Urea fertilizer	12	48%
10	Phonska fertilizer	7	28%
11	NPK fertilizer	1	4%
12	SS fertilizer	3	12%
13	ZA fertilizer	5	20%
14	TSP fertilizer	1	4%
15	Fertilizer ST	3	12%
16	Gramoxon pesticide	6	24%
17	Basmilang Pesticides	7	28%
18	Pesticide Push up	4	16%

Table 4 shows the selected data items with a minimum support of 16%. Table 4. Types of set items that meet a predetermined minimum support.

Table 4. Types of set items that meet a predetermined minimum support

No	Item Name	Number of Items	Support
1	Corn Seed	9	36%
2	Kale seeds	4	16%
3	Urea fertilizer	12	48%
4	Phonska fertilizer	7	28%
5	ZA fertilizer	5	20%
6	Gramoxon pesticide	6	24%
7	Basmilang Pesticides	7	28%
8	Pesticide Push up	4	16%

b) The Iteration-2 formation is formed from product items that meet minimum support by combining all items into two combinations. The results of the combination of 2 items can be seen in table 5 below:

Table 5. Combination of 2 itemset

No	Two Item Names	amount
1	Corn Seeds, Kale Seeds	3
2	Corn Seeds, Urea Fertilizer	4
3	Corn Seeds, Phonska Fertilizer	1
4	Corn Seeds, ZA Fertilizer	1
5	Corn Seeds, Gramoxon Pesticides	2
6	Corn Seeds, Basmilang Pesticides	1
7	Corn Seed, Pesticide Push Up	1
8	Kale seeds, Urea fertilizer	1
9	Water spinach seeds, Phonska fertilizer	0
10	Water spinach seeds, ZA fertilizer	0
11	Kale seeds, Gramoxon Pesticide	1
12	Kale seeds, Basmilang pesticide	0
13	Kale seeds, Pesticide Push up	0
14	Urea Fertilizer, Phonska Fertilizer	4
15	Urea Fertilizer, ZA Fertilizer	2
16	Urea Fertilizer, Gramoxon Pesticides	2
17	Urea Fertilizer, Basmilang Pesticides	4
18	Urea Fertilizer, Pesticide Push up	2
19	Phonska Fertilizer, ZA Fertilizer	1
20	Phonska Fertilizer, Gramoxon Pesticide	2
21	Phonska Fertilizer, Basmilang Pesticides	4
22	Phonska Fertilizer, Pesticide Push up	2
23	ZA Fertilizer, Gramoxon Pesticides	0
24	ZA fertilizer, Basmilang pesticides	1
25	ZA Fertilizer, Pesticide Push up	1
26	Gramoxon Pesticides, Basmilang Pesticides	2
27	Gramoxon Pesticides, Push up Pesticides	2
28	Pesticides Basmilang, Pesticides Push up	0

Finding the minimum value of 2 items can be obtained by the following formula:
Support (A,B) = P (A ∩ B)

$$Support\ A = \frac{\sum Jumlah\ Transaksi\ Mengandung\ A\ dan\ B}{\sum Total\ Transaksi} \quad (2)$$

The result of the 2nd iteration calculation.

Table 6. Combination of 2 itemset and Support value

No	Two Item Names	amount	Support
1	Corn Seeds, Kale Seeds	3	12%
2	Corn Seeds, Urea Fertilizer	4	16%
3	Corn Seeds, Phonska Fertilizer	1	4%
4	Corn Seeds, ZA Fertilizer	1	4%
5	Corn Seeds, Gramoxon Pesticides	2	8%

No	Two Item Names	amount	Support
6	Corn Seeds, Basmilang Pesticides	1	4%
7	Corn Seed, Pesticide Push Up	1	4%
8	Kale seeds, Urea fertilizer	1	4%
9	Water spinach seeds, Phonska fertilizer	0	0%
10	Water spinach seeds, ZA fertilizer	0	0%
11	Kale seeds, Gramoxon Pesticide	1	4%
12	Kale seeds, Basmilang pesticide	0	0%
13	Kale seeds, Pesticide Push up	0	0%
14	Urea Fertilizer, Phonska Fertilizer	4	16%
15	Urea Fertilizer, ZA Fertilizer	2	8%
16	Urea Fertilizer, Gramoxon Pesticides	2	8%
17	Urea Fertilizer, Basmilang Pesticides	4	16%
18	Urea Fertilizer, Pesticide Push up	2	8%
19	Phonska Fertilizer, ZA Fertilizer	1	4%
20	Phonska Fertilizer, Gramoxon Pesticide	2	8%
21	Phonska Fertilizer, Basmilang Pesticides	4	16%
22	Phonska Fertilizer, Pesticide Push up	2	8%
23	ZA Fertilizer, Gramoxon Pesticides	0	0%
24	ZA fertilizer, Basmilang pesticides	1	4%
25	ZA Fertilizer, Pesticide Push up	1	4%
26	Gramoxon Pesticides, Basmilang Pesticides	2	8%
27	Gramoxon Pesticides, Push up Pesticides	2	8%
28	Pesticides Basmilang, Pesticides Push up	0	0%

The 2nd iteration has found 2 set items that meet the 16% minimum support.

Table 7. Combination of 2 itemset and Support value

No	Two Item Names	amount	Support
1	Corn Seeds, Urea Fertilizer	4	16%
2	Urea Fertilizer, Phonska Fertilizer	4	16%
3	Urea Fertilizer, Basmilang Pesticides	4	16%
4	Phonska Fertilizer, Basmilang Pesticides	4	16%

c) Rules of Association

In the association rules, this is done after a high-frequency pattern is found, which then looks for the association rules first looking for a certainty value or commonly referred to as the confidence value, which functions to see how close the relationship is between these items, first determine the confidence value of 50% and the rules can be found. Association by applying the following formula.

$$Confidence = P(A|B) \frac{\sum Jumlah\ Transaksi\ Mengandung\ A\ dan\ B}{\sum Total\ Transaksi\ A} \quad (3)$$

Table 8. Value of Association Rules with Confidence

Item Combinations	Confident	
If you buy Corn Seeds, you will buy Urea Fertilizer	9/9	100%
If you buy Urea Fertilizer, you will buy Corn Seed Sticks	9/12	75,0%
If you buy Urea, you will buy Phonska Fertilizer	7/12	58%
If you buy Phonska Fertilizer, you will buy Urea Fertilizer	7/7	100%

If you buy Urea Fertilizer, you will buy Basmilang Pesticides	7/12	58%
If you buy Basmilang Pesticides, you will buy Urea Fertilizer	7/7	100%
If you buy Phonska Fertilizer, you will buy Basmilang Pesticides	7/7	100%
If you buy Basmilang Pesticides, you will buy Phonska Fertilizer	7/7	100%

By setting a minimum confidence value of 80%, the results of a rule that meet the minimum confident value can be seen in table 9 below:

Table 9. Value of Association Rules with Confidence

Item Combinations	Support	Confident
If you buy Corn Seeds, you will buy Urea Fertilizer	16%	100%
If you buy Phonska Fertilizer, you will buy Urea Fertilizer	16%	100%
If you buy Basmilang Pesticides, you will buy Urea Fertilizer	16%	100%
If you buy Phonska Fertilizer, you will buy Basmilang Pesticides	16%	100%
If you buy Basmilang Pesticides, you will buy Phonska Fertilizer	16%	100%

4. Conclusion

On the basis of the discussion with the Apriori algorithm, a number of conclusions can be drawn, namely the use of data mining, by using data on the sale of agricultural equipment, because it can establish a relationship between one item and another, so that it can be used as valuable information to set patterns. The placement of goods and the preparation of stock of what types of goods are needed later.

References

[1] N. F. Hasan, R. Hammad, D. E. Profesi, and K. Kusrini, "Sistem Pendukung Keputusan Penentuan Kombinasi Paket Produk Pertanian Menggunakan Algoritma Apriori," *J. Eksplora Inform.*, vol. 9, no. 1, pp. 38–49, 2019.

[2] M. P. Tana, F. Marisa, and I. D. Wijaya, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Pada Toko Oase Menggunakan Algoritma Apriori," *JIMP-Jurnal Inform. Merdeka Pasuruan*, vol. 3, no. 2, 2018.

[3] A. Waluyo, H. Jatnika, M. R. S. Permatasari, T. Tuslaela, I. Purnamasari, and A. P. Windarto, "Data Mining Optimization uses C4.5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, pp. 1–9, 2020.

[4] F. Rahman, I. I. Ridho, M. Muflih, S. Pratama, M. R. Raharjo, and A. P. Windarto, "Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 835, no. 1, 2020.
[5] S. Mujanah, I. Wahyurini, and Murgianto, "The effect of perceived organizational support, self-efficacy, emotional intelligence on employee performance with interpersonal communication as intervening variable," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, no. 5, pp. 612–622, 2019.
[6] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, "Classification of natural disaster prone areas in Indonesia using K-means," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 8, pp. 87–98, 2018.
[7] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 855, 2020.
[8] A. P. Windarto, J. Na, and A. Wanto, "Bagian 2 : Model Arsitektur Neural Network dengan Kombinasi K- Medoids dan Backpropagation pada kasus Pandemi COVID-19 di Indonesia," vol. 4, pp. 1175–1180, 2020.
[9] J. Hong, R. Tamakloe, and D. Park, "Discovering Insightful Rules among Truck Crash Characteristics using Apriori Algorithm," *J. Adv. Transp.*, vol. 2020, 2020.
[10] S. Raj, D. Ramesh, and K. K. Sethi, "A Spark-based Apriori algorithm with reduced shuffle overhead," *J. Supercomput.*, vol. 77, no. 1, pp. 133–151, 2021.
[11] X. Yuan, "An improved Apriori algorithm for mining association rules," *AIP Conf. Proc.*, vol. 1820, no. March 2017, 2017.
[12] A. Soni, A. Saxena, and P. Bajaj, "A methodological approach for mining the user requirements using apriori algorithm," *J. Cases Inf. Technol.*, vol. 22, no. 4, pp. 1–30, 2020.
[13] N. Verma, D. Malhotra, and J. Singh, "Big data analytics for retail industry using MapReduce-Apriori framework," *J. Manag. Anal.*, vol. 7, no. 3, pp. 424–442, 2020.
[14] T. Kristiana, S. A. Putri, Nurmalasari, R. I. Handayani, N. Merlina, and N. Yunita, "Association Rule Implementation Using Algorithm Apriori to Analize Fishing Pattern in Indonesia," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020.
[15] H. Jantan and A. Z. M. Jamil, "Association rule mining based crime analysis using apriori algorithm," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.5 Special Issue, pp. 18–24, 2019.
[16] X. Yang, X. Lin, and X. Lin, "Application of Apriori and FP-growth algorithms in soft examination data analysis," *J. Intell. Fuzzy Syst.*, vol. 37, no. 1, pp. 425–432, 2019.
[17] R. D. Yulanda, S. Wahyuningsih, and F. D. T. Amijaya, "Association rules with apriori algorithm and hash-based algorithm," *J. Phys. Conf. Ser.*, vol. 1277, no. 1, 2019.
[18] K. M. Anil Kumar, B. Ajay, R. Shashank, and D. A. Amogha Subramanya, "An apriori method for topic extraction from text files," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 2516–2521, 2019.
[19] L. Liu, S. Yu, X. Wei, and Z. Ning, "An improved Apriori–based algorithm for friends recommendation in microblog," *Int. J. Commun. Syst.*, vol. 31, no. 2, pp. 1–11, 2018.

Rulinawaty

ORIGINALITY REPORT

6%	5%	7%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Institut Teknologi Nasional Malang Student Paper	4%
2	Yusuf Kurnia, Yohanes Isharianto, Yo Ceng Giap, Aditiya Hermawan, Riki. "Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm", Journal of Physics: Conference Series, 2019 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		