PAPER • OPEN ACCESS

Data Mining Technique to Determine the Pattern of Fruits Sales & Supplies Using Apriori Algorithm

To cite this article: Eni Heni Hermaliani et al 2020 J. Phys.: Conf. Ser. 1641 012070

View the article online for updates and enhancements.



This content was downloaded from IP address 115.124.73.190 on 23/03/2021 at 08:42

Data Mining Technique to Determine the Pattern of Fruits Sales & Supplies Using Apriori Algorithm

Eni Heni Hermaliani ¹*, Laela Kurniawati², Tuti Haryanti ³, Nisa Mutiah⁴, Aan Kurniawan⁵, and Bahrun Said Renhoran⁶

^{1,5,6}, Computer Science, STMIK Nusa Mandiri, Jakarta, 10420, Indonesia ^{2,3,4}, Information System, STMIK Nusa Mandiri, Jakarta, 10420, Indonesia

E-mail: enie_h@nusamandiri.ac.id

Abstract. Advances in technological developments in the current 4.0 era require retail businesses to increase sales and develop marketing strategies. Determination of which products are widely sold and which products will be propagated in inventory is very important for retail businesses to prevent data accumulation. Data Mining has been widely used to conduct analysis, determine patterns and associations. In this paper, we propose a basic methodology for calculating association analysis with apriori algorithms used to process the most sold products and which products will be propagated in the sales inventory. The results of this study are the calculation model of high-frequency pattern analysis and the formation of association rules and item set combination patterns resulting from the sale of fruit products from retail with the highest support and confidence of the most sold fruit products. Therefore the apriori algorithm can help develop sales strategies.

Keywords : Determining the pattern of fruits Sales, Apriori Algorithm, Data Mining Technique

1. Introduction

The progress of the development of information technology is currently giving direction for retail companies' business of product sales in a time of increasingly fierce market competition to meet customer desires. Companies need to dig up valuable and detailed information, configure goods to sell better, and improve the efficiency of market operations. At present retail business can collect product data and customer data in a very large database by implementing data mining processes through pattern analysis and data search with associations that aim to provide the best service to consumers. Big data is modeled, selected, and explored to obtain information that can be understood and useful for humans [1]. Data mining (DM) is a promising and up-to-date field as part of computer science, where the use of data mining has become a top priority in the information industry and society in recent years. The existence of big data, it is very important to do the right use in extracting hidden knowledge in the data warehouse, data mart, or repository [2], [3]. The extraction of useful information on a large amount of data in a database can use a technique that is owned by DM [4], [5]. Association rules as a result of applying the DM algorithm can be used to find relationships between sales items that are different from big data analysis [6]. Increased efficiency of extraction rules in association rules can be used with a hierarchical method through a hybrid approach that combines objective and subjective actions [7]. Apriori algorithm is one of the most popular algorithms and as part of unsupervised learning for mining association rules [8]. To identify the set of product data items



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd

that are often purchased and collected in a database can be mining using apriori algorithm [9]. In previous research to increase retail business sales, association rules were used which provided various benefits to retailers by analyzing customer purchasing patterns, obtaining timely and accurate product pipeline status, and inventory. The source of the transaction database in data mining and extracting significant knowledge comes from the Point of Sales data [10]. While research [11] on product sales data has found the rules of association between products after analyzing data based on shopping records in supermarkets as a dataset in the form of data sets derived from transactions at the checkout with a priori algorithm found product data sets related to each other, set layout shelves according to product relationships and increase sales and sales revenue. Strengthened by [12] in his research that to determine associations according to lovalty, apriori algorithms can be applied to the set of transactions of customer groups to find associations between products. Information sourced from large amounts of transaction data with association rule mining can be found patterns, associations and relationships in datasets that help retailers identify relationships between items that consumers often buy through the implementation of a priori algorithms that can be found frequently [13]. Supported by the results of [14] the calculation and discovery of association rules with good results have been proven by applying apriori algorithm as an innovative method, in this case, using two parameters in the form of a support value and a minimum trust value and [15]

This study aims to determine the extent to which apriori algorithms can help the development of marketing strategies, obtain the association pattern, and determine the best-selling products. In this paper, we conduct sales analysis using association rules to find attributes that appear at one time and find solutions regarding frequency patterns and determine association rules to find models based on the best and best confidence and support. The apriori algorithm used with the experimental dataset is processed sourced from a retail company in the department of fresh vegetables and fruit. Whereas for the calculation of the algorithm the Tanagra application is used as a data mining tool in analyzing data exploration, statistical learning, machine learning, and database areas.

2. Method

The apriori algorithm method is used to be able to find the attributes that appear at one time. The application of the basic methodology of association analysis according to the steps in apriori algorithm calculation is divided into two stages, namely:

a. High-Frequency Pattern Analysis

Look for a combination of items that meet the minimum requirements and an item's support value is obtained.

b. Formation of Association Rules

After all, high-frequency patterns are found, next look for association rules that meet the minimum requirements for confidence.

Then to determine the rules of the association to be chosen it must be sorted by support and confidence. n rules are taken as many as the rules that have the greatest results. This research has used datasets sourced from data on fruit product sales at retail companies for one year, which has 150 types of fruit items with a total number of sales transactions per month around 1500 - 2500 Qty. This research applies the proposed step framework, where the entire composition of the composition can be seen in Figure 1.



Figure 1. Method used

3. Result and Discussion

At this stage analysis of sales data for fruit products in retail companies for 1 year with transactions obtained from monthly sales for top sales based on monthly reports as in table 1.

Table 1. Data on Fruits that are Sold

Itemset	The fruits name
SP	CAVENDISH BANANA SUNPRIDE KG
SM	NON-SEEDED WATERMELON KG
PC	ORGANIC CALIFORNIA PAPAYA KG
LL	LONGAN LOKAL KG
JB	IMLEK HONEY BALI ORANGE
BL	STAR FRUIT KG
$_{\rm JP}$	PONKAM MANDARIN ORANGE KG
JK	KINNO MANDARIN ORANGE KG
\mathbf{PF}	PEARS FORELL KG
\mathbf{PX}	PEAR XIANG LIE KG

Based on data from the sale of fruit for 1 year, the transaction pattern is obtained by analyzing the 3 most fruit names that are sold every month. The names of the most sold fruits every month can be seen in table 2: a. Making a Tabular Format

Table 2.	Pattern	Fruits	Sales	of	Transaction	in	2018
Labic 2.	1 autorn	1 I UI UB	Daros	or	ransaction	111	2010

Months	Itemset
1	SP,SM,PC
2	SP,PC,SM
3	SP,PC,SM
4	LL,SP, JB
5	SP,JB,BL
6	SP, JP, PC
7	SP,JK,PF
8	SP,PX,JB
9	SP,SM,JK
10	SP,PC,SM
11	SP,PC,SM
12	SP, JB, PC

The tabular format was created with the aim of making it easier to do association analysis. Namely in the form of a binary representation table for each transaction, the attribute value is expressed as 1 if a transaction contains a purchase for the item in question, and the attribute value is 0 if in a transaction there is no purchase of the item in question. 150 types of fruit have been tested with the first-order results are Cavendish Banana Sunpride KG, the second sequence is Organic California Papaya KG, and the third sequence is Non-Seeded Watermelon KG.

b. Itemset Formation

The following is a settlement with case examples based on the data that has been provided. The process of forming C1 or referred to as 1 itemset with a minimum amount of support =

30%. The results of the formation of 1 itemset with minimum support of 30%, can be seen in table 4. Known that meet the minimum support standards are Cavendish Banana Sunpride KG 100%, Non-Seeded Watermelon KG 41.7%, Organic California Papaya KG 50%, Imlek Honey Bali Orange 33.4%.

Table 3.	Support	of each	itemset
----------	---------	---------	---------

Itemset	Support
SP	100%
\mathbf{SM}	41,7%
\mathbf{PC}	50%
$_{\mathrm{JB}}$	$33,\!4\%$

Judging from the results of 1 itemset can meet the maximum support requirement of 30%, then the calculation is continued for the combination of 2 itemsets. Formation of 2 itemsets is not done on all itemsets, but only on itemsets that have supported greater than the minimum support.

c. 2 Itemsets Combination

The process of forming C2 or referred to as 2 itemsets with a minimum amount of support of 30%. The result of the combination of 2 itemsets that can meet the minimum support requirement is 30%, then the combination of 2 itemsets that does not meet the minimum support will be eliminated, can be seen as in table 5.

d. 3 Itemsets Combination

Itemset	Support
SP, SM	41,7%
SP, PC	50%
SP, JB	$33,\!4\%$
SM, PC	$33,\!4\%$

Table 4. Minimum Support and 2 Itemsets 30	%
--	---

The process of forming C3 or referred to as 3 itemsets with a minimum amount of support of 30%. Based on the results of a combination of 3 itemsets that can meet the maximum support requirement of 30%, then the combination of 3 itemsets that does not meet the minimum supply will be eliminated, can be seen as table 6 and can be known the combination of 3 itemsets that can meet the minimum support requirement of 30% is Cavendish Banana Sunpride KG, Non-Seeded Watermelon KG, and Organic California Papaya KG 33.4%

Table 5. Support of 3 combinations of 3 itemsets

Itemset	Support
SP,SM,PC	$33,\!4\%$

e. Formation of Association Rule

Based on the calculation of the combination of 2 itemsets, item sets that have high-frequency pattern results that meet the requirements can be seen in table 7.

Table 6. Results of the high-frequency pattern combination of 2 itemsets

Itemset	Support
SP,SM	33,4%
SP,PC	50%
SP,JB	$33,\!4\%$

Because the 3 itemsets that meet the minimum support requirement is only 1 itemset, the confidence value that is sought is 2 itemset combinations. After all high-frequency patterns have been found, then look for the association rules that meet the minimum requirements for confidence by calculating confidence or associative $A \rightarrow B$, with a minimum of 50% confidence. Based on the process of calculating a minimum value of 50% confidence, 2 itemsets that meets the minimum requirement of 50% confidence are Cavendish Banana Sunpride KG and Non-Seeded Watermelon KG 50% and Cavendish Banana Sunpride KG and Organic California Papaya KG, as shown in table 8.

Table 7. Confidence from 2 itemsets

Rules	Confidence
If you buy SP, you will buy SM	5/12 50%
If you buy SP, you will buy PC	6/12 50%

f. Final Association Rules

The final association rules are ordered based on a minimum of 30% support and minimum confidence of 50% that, has been determined, can be seen in table 9.

In table 9, it can be seen that if the buyer buys Cavendish Banana Sunpride KG, he will buy

Rules	Support	Confidence
If you buy SP, you will buy SM	$33,\!4\%$	50%
If you buy SP, you will buy PC	50%	50%

 Table 8. Final Association Rules

Non-Seeded Watermelon KG, has a confidence value of 50%, besides that, if the buyer buys Cavendish Banana Sunpride KG then will buy the Organic California Papaya KG by having a confidence value of 50%.

4. Conclusion and Future Works

Data mining using a priori algorithm can generate new information patterns to determine sales patterns in the rules of the goods association and develop marketing strategies to market other

products, by examining what are the advantages of the most sold products. Through the Tanagra tools, support and confidence are formed which will then produce a final association that meets support and confidence. The sale of the most sold fruit products to retail business is known by using a priori algorithm, by looking at products that meet a minimum of support and minimum confidence. Thus the most sold products are Cavendish Banana Sunpride KG, Non-Seeded Watermelon KG, Organic California Papaya KG. The results of the final association rules are known if buying Sunpride Banana Cavendish KG and Non-Seeded Watermelon KG , they will buy Organic California Papaya KG with 33.33% support and 80.00% confidence. If you buy Non-Seeded Watermelon KG, you will buy Cavendish Banana Sunpride KG and Organic California Papaya KG with 33.33% support and 80.00% confidence. If you buy Non-Seeded Watermelon KG, you will buy Cavendish Banana Sunpride KG and Organic California Papaya KG with 33.33% support and 80.00% confidence. If you buy Non-Seeded Watermelon KG, you will buy Organic California Papaya KG with 33.33% support and 80.00% confidence. For further research, there are some suggestions that we propose, 1) Can use even more data and a variety of products in sales transactions; 2) Can be compared with other algorithms, namely the FP-Growth algorithm, the Tertius algorithm, the Generalized Sequential Pattern (GSP) algorithm in finding more accurate results.

Acknowledgement

The authors would like to thank to retail business for granting research data permission in this study. Thank you also to all parties for your support until this paper has been processed and published.

References

- V. Singh and K. Kumar, "Data Mining and Knowledge Management," Int. Res. J. Eng. Technol., vol. 4, no. No.2, February 2017, pp. 200–206, 2017.
- [2] J. Han, J. Pei, and M. Kamber, "Data mining: concepts and techniques," Elsevier, vol. 2, 2011.
- [3] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA," Indones. J. Electr. Eng. Comput. Sci., vol. 9, no. No.2, February 2018, pp. 447–459, 2018.
- [4] F. M. and H. M. H. Ba-Alwi, "Comparative Study for Analysis the Prognostic in Hepatitis Data: Data Mining Approach," Int. J. Sci. Eng. Res., vol. 4, no. 8, p. 64, 2013.
- [5] U. Fayyad, P. G. Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," American Association for Artificial Intelligence Magazine, p. 18, 1996.
- [6] F. . Ba-Alwi, "Discovery of novel association rules based on genetic algorithms," Br. J. Math. Comput. Sci., vol. 4, no. 23, p. 17, 2014.
- [7] M. Z. Sušac and Adela, "Discovering market basket patterns using hierarchical association rules," no. 31 October, 2015.
- [8] A. K. and R. N. P. Shrivastava, "Implementation of Apriori Algorithm using WEKA," vol. 1, no. 1, p. 4, 2014.
- [9] I. R. V. Srinivasa Kumar, R.Renganathan, C.VijayaBanu, "Consumer Buying Pattern Analysis using Apriori Association Rule," Int. J. Pure Appl. Math., vol. 119, no. 7, 2018.
- [10] N. Verma, D. Malhotra, and S. Jatinder, "Big data analytics for retail industry using MapReduce-Apriori framework," J. Manag. Anal., 2020.
- [11] P. Yazgan, "ASSOCIATION RULES AND MARKET BASKET ANALYSIS: A CASE STUDY IN RETAIL SECTOR," Istanbul Commerce University, 2016.
- [12] J. Silva et al., "Association Rule Mining for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm," in ICACDS 2019, 2019, pp. 488–497.
- [13] M. Kavitha and S. Subbaiah, "Association Rule Mining using Apriori Algorithm for Extracting Product Sales Patterns in Groceries," Int. J. Eng. Res. Technol., vol. 08, no. 03, pp. 1–4, 2020.
- [14] Y. Kurnia, Y. Isharianto, Y. C. Giap, A. Hermawan, and 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," in 1st International Conference on Advance and Scientific Innovation (ICASI) - IOP Conf. Series: Journal of Physics: Conf. Series 1175, 2019, pp. 1–6.
- [15] R. Husna, R. Lestari, and Y. Hendra, "Inventory model of goods availability with apriori algorithm," in ICOMSET, IOP Conf. Series: Journal of Physics: Conf. Series 1317, 2018, pp. 1–8.