Utilizing Apriori Data Mining Techniques on Sales Transactions

Umbas Krisnanto
Fakultas Ekonomi Dan Bisnis, Perbanas Institut, Jakarta, Indonesia.
E-mail: umbas@perbanas.id

J. Juharsah
Universitas Halu Oleo, Indonesia.
E-mail: juharsah2005@gmail.com

Purnama Putra
Universitas Islam, Bekasi, Indonesia.
E-mail: putra.purnama@unismabekasi.ac.id

Andini Dani Achmad
Universitas Hasanuddin, Makassar, Indonesia.
E-mail: andini.achmad@gmail.com

Elkana Timotius
Universitas Kristen Krida Wacana, Indonesia.
E-mail: mr.elkana@gmail.com

Received October 08, 2021; Accepted December 25, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V19I1/WEB19376

Abstract

The establishment of a marketing strategy is important for every business actor in the competitive world of business. Business operators must be able to develop sound marketing strategies to influence the attractiveness of consumers and to buy interest in the products provided so that the enterprise they operate can compete and have a market share and to maximize sales sales. To implement marketing strategies, references are required so that promotions can reach the right target, for example by seeking similarities between items. By using data mining techniques, these studies apply the a priori approach to the promotion of customer product recommendations by association rules on product sales transaction datasets to aid in the formation of applications between product items. The dataset represents a sample of sales of products for 2020. The application used for analyzing is RapidMiner, where a support value of > 20% and confidence of > 60% is determined. Each product package promoted is made up of 2 products from the calculation results. The two best rules that have value confidence is combined with 2 items (Cre1→Cre2), (Cre1→Cre12) and (Cre9→Cre10). Based on the minimum support and confidence values that have been set, the results of the a

priori method can produce association rules that can be used as a reference in product promotion and decision support in providing product recommendations to consumers.

Keywords

Promotion, Rule Association, A Priori, Marketing Strategy.

Introduction

Marketing strategy is critical for business actors in today's competitive business world [1]. Business actors must be able to implement good marketing strategies in order to influence consumers' attractiveness and purchasing interest in the products offered [2], [3]. Furthermore, the business can compete while maintaining market share, allowing for increased sales and maximum profit. One possible strategy is to promote [4], [5]. The use of promotion is a step that can be taken to introduce and inform consumers about products sold. People require practicality in meeting their needs in a variety of ways, including the provision of household appliances and furniture, as the times progress. There are numerous stores that sell household furniture. Every furniture store has sales transaction data. Sales transaction data is one type of information that can be used to make business decisions. The majority of sales transaction data is not reused and is only saved as an archive when creating a report. Whereas transaction data can be used to create a business policy, specifically by incorporating the information into a business strategy [6], [7].

Data mining is [8], [9] one of the sciences that can be used in situations like this. Unused sales transactions can be extracted and reprocessed into useful information using data mining techniques. This data mining technique uses historical data to identify frequent patterns that can be used to benefit these marketing activities. Frequent pattern data contains information on patterns that frequently appear in transaction data. The use of association rules as a data mining method that can identify similarities between items is one technique for exploring Frequent pattern information. The a priori method [10], [11] can be used to generate candidate item combinations for applying associative rules with a specific frequency value. This method has been used in a variety of applications. As an example, consider the research conducted [12] on the security system. The a priori method, which is the classic association rule in a web-based intrusion detection system, is proposed in this paper, and the rule base generated by the a priori method is used to identify various attacks. As a result, association rules and priori methods can be used to solve Intrusion Detection problems. The research was then conducted [13] on efforts to reduce the severity of traffic accidents by examining the condition of data obtained using the association rules method with the a priori calculation method. The sought-after

relationship pattern is the relationship between itemset variables involved in accidents involving four variables that describe the identity of the perpetrator, namely gender, age, profession, and level of education, as well as 22 attribute datasets. As a result [14], the significance of association rules can be seen through two parameters: minimum support (percentage of item combinations from all transactions) and minimum confidence (strong relationship between items in associative rules). In contrast to the use of the a priori algorithm in the preceding system, the purpose of this study is to investigate the use of this algorithm and association rule on the sales transaction dataset to provide promotional product recommendations to customers. In 2020, the dataset is in the form of sales transaction data. The highest frequency pattern analysis yields a set of items that meet the specified minimum support requirements. The formation of the association rule pattern is carried out in order to find association rules that meet the minimum confidence of the found high frequencies.

Research Methodology

The a priori algorithm is applied to a dataset of transaction history data in this study. Data collection, data pre-processing, analysis of the highest frequency pattern using a priori algorithms, formation of association rule patterns, and testing of experimental results are the stages of the research. The use of association rules with a priori algorithms has benefits in terms of simplicity and the ability to handle large amounts of data, mh limited data processing capabilities. The RapidMiner application is used to perform analysis in this study.

The Apriori Method works as follows [15]:

1) Determine the bare minimum of assistance.
2) Iteration 1: count items from support (transactions containing all items) by scanning the database for 1-itemset, after 1-itemset is obtained, is it above the minimum support from 1-itemset, if so, 1-itemset will be a high frequent pattern.
3) Iteration 2: To get the 2-itemset, combine the previous k-itemset, then scan the database again to count the items with support. The most popular itemset will be chosen as the candidate's most frequent pattern.
4) Set the k-itemset value from the support that has met the k-minimum itemset's support
5) Repeat the process for the next iteration until no more k-itemsets meet the minimum support.

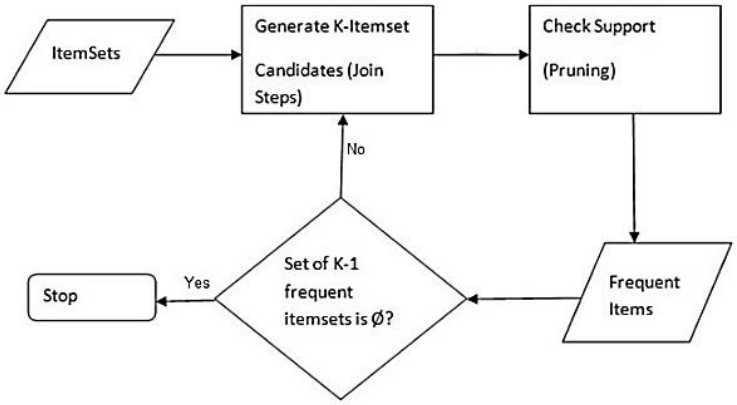

Figure 1 Flowchart illustration of the Apriori method

Support and Confidence for Itemset A and B are represented by formulas[16]:

$$\text{Support (A)} = \frac{\text{Number of transaction in which A appears}}{\text{Total number of transactions}} \quad (1)$$

$$\text{Confidence (A}\rightarrow\text{B)} = \frac{\text{Support(AUB)}}{\text{Support(A)}} \quad (2)$$

Results and Discussion

The information was gathered through observation and interviews with the owner of Lido Gallery. A sample of product sales transactions in 2020 was used as the dataset. Based on research, if the research subject is less than 100, it can be used entirely; if the research subject is more than 100, it can be used 10%-15% or 20%-25% or more [17]. As shown in Table 1, the following is a sample dataset of sales transactions for 2020.

Table 1 Sales Transaction Data at Lido Gallery in 2020

No	Furniture for sale
1	Wardrobe, Dressing Table, Springbed, Buffet
2	Sofa, Sideboard, Television, Air Conditioner
3	Kitchen Set, Dish Cabinet, SofaSo
4	Dressing Table, Wardrobe, Air Conditioner, Washing Machine
5	Television, Louspeaker, Buffet
6	Kitchen Set, Refrigerator, Electric Oven
7	Electric Oven, Washing Machine, Buffet
8	Sofa, Springbed, Wardrobe, Buffet
9	Springbed, Wardrobe, Sofa, Kitchen Set
10	Refrigerator, Kitchen Set, Television
11	Wardrobe, Dressing Table, Television, Sofa
12	Springbed, Television, Dressing Table, Air Conditioner, Refrigerator
13	Wardrobe, Dressing Table, Television, Sofa, Springbed
14	Sideboard, Air Conditioner, Refrigerator, Kitchen Set, Washing Machine

Source: processed data

Table 1 contains several transactions, each of which contains several items purchased by consumers. Each item will be coded as follows to make it easier to find rules:

Table 2 Product Data Assumptions

Item Set	Code
Wardrobe	Cre1
Dressing table	Cre2
Television	Cre3
Sideboard	Cre4
Cabinet for plates	Cre5
Air conditioning	Cre6
Loudspeaker	Cre7
Washing machine	Cre8
Refrigerator	Cre9
Kitchen set	Cre10
Sofa	Cre11
Spring bed	Cre12
Electric oven	Cre13

Based on the transaction data in table 2, as a first step to find the support value of each item. Table 3 shown is the formation of 1 itemset with its frequency and support values. The support value is obtained by forming 1 itemset that meets the specified minimum support (minsup) requirements.

Table 3 Support Value

Item Set	Code	Support Account	Support	Support
Wardrobe	Cre1	6	=6/14*100%	43%
Dressing table	Cre2	5	=5/14*100%	36%
Television	Cre3	6	=6/14*100%	43%
Sideboard	Cre4	4	=4/14*100%	29%
Cabinet for plates	Cre5	1	=1/14*100%	7%
Air conditioning	Cre6	4	=4/14*100%	29%
Loudspeaker	Cre7	1	=1/14*100%	7%
Washing machine	Cre8	3	=3/14*100%	21%
Refrigerator	Cre9	4	=4/14*100%	29%
Kitchen set	Cre10	4	=4/14*100%	29%
Sofa	Cre11	6	=6/14*100%	43%
Spring bed	Cre12	5	=5/14*100%	36%
Electric oven	Cre13	2	=2/14*100%	14%

Then, using Table 3, combine each item into two items, as shown in Table 4:

Table 4 Item Combinations

Code (item)	Suppot Account	Code (item)	Suppot Account
Cre1, Cre2	4	Cre4, Cre11	2
Cre1, Cre3	1	Cre4, Cre12	2
Cre1, Cre4	2	Cre4, Cre13	1
Cre1, Cre5	0	Cre5, Cre6	0
Cre1, Cre6	1	Cre5, Cre7	0
Cre1, Cre7	0	Cre5, Cre8	0
Cre1, Cre8	0	Cre5, Cre9	0
Cre1, Cre9	0	Cre5, Cre10	1
Cre1, Cre10	1	Cre5, Cre11	1
Cre1, Cre11	3	Cre5, Cre12	0
Cre1, Cre12	4	Cre5, Cre13	0
Cre1, Cre13	0	Cre6, Cre7	0
Cre2, Cre3	3	Cre6, Cre8	2
Cre2, Cre4	1	Cre6, Cre9	2
Cre2, Cre5	0	Cre6, Cre10	1
Cre2, Cre6	2	Cre6, Cre11	1
Cre2, Cre7	0	Cre6, Cre12	1
Cre2, Cre8	1	Cre6, Cre13	0
Cre2, Cre9	1	Cre7, Cre8	0
Cre2, Cre10	0	Cre7, Cre9	0
Cre2, Cre11	2	Cre7, Cre10	0
Cre2, Cre12	3	Cre7, Cre11	0
Cre2, Cre13	0	Cre7, Cre12	0
Cre3, Cre4	2	Cre7, Cre13	0
Cre3, Cre5	0	Cre8, Cre9	1
Cre3, Cre6	2	Cre8, Cre10	1
Cre3, Cre7	1	Cre8, Cre11	0
Cre3, Cre8	0	Cre8, Cre12	0
Cre3, Cre9	2	Cre8, Cre13	1
Cre3, Cre10	1	Cre9, Cre10	3
Cre3, Cre11	3	Cre9, Cre11	0
Cre3, Cre12	2	Cre9, Cre12	1
Cre3, Cre13	0	Cre9, Cre13	1
Cre4, Cre5	0	Cre10, Cre11	1
Cre4, Cre6	2	Cre10, Cre12	0
Cre4, Cre7	0	Cre10, Cre13	0
Cre4, Cre8	1	Cre11, Cre12	3
Cre4, Cre9	1	Cre11, Cre13	0
Cre4, Cre10	1	Cre12, Cre13	0

The steps that follow determine the frequent pattern and search for the support value of combination 2 or A U B (A Combined B). The process of forming two itemsets is performed on itemsets that meet the minimum support requirements shown in Table 5.

Table 5 Frequency Pattern

2 Item Combination Pattern	Support Account	Support A U B
Cre1, Cre2	4	29%
Cre1, Cre4	2	14%
Cre1, Cre11	3	21%
Cre1, Cre12	4	29%
Cre2, Cre3	3	21%
Cre2, Cre6	2	14%
Cre2, Cre11	2	14%
Cre2, Cre12	3	21%
Cre3, Cre4	2	14%
Cre3, Cre6	2	14%
Cre3, Cre9	2	14%
Cre3, Cre11	3	21%
Cre3, Cre12	2	14%
Cre4, Cre11	2	14%
Cre4, Cre12	2	14%
Cre6, Cre8	2	14%
Cre6, Cre9	2	14%
Cre9, Cre10	3	21%
Cre11, Cre12	3	21%

The next step is to create association rules. Before you start, calculate the confidence value of each item in the frequency pattern table shown in Table 6 below.

Table 6 Combination of 2 Itemsets with Confidence Value

2 Itemset Combination Pattern	*Confidence*	
Cre1→ Cre2	4/6*100%	66%
Cre1→ Cre4	2/6*100%	33%
Cre1→ Cre11	3/6*100%	50%
Cre1→ Cre12	4/6*100%	66%
Cre2→ Cre3	3/5*100%	60%
Cre2→ Cre6	2/5*100%	40%
Cre2→ Cre11	2/5*100%	40%
Cre2→ Cre12	3/5*100%	60%
Cre3→ Cre4	2/6*100%	33%
Cre3→ Cre6	2/6*100%	33%
Cre3→ Cre9	2/6*100%	33%
Cre3→ Cre11	3/6*100%	50%
Cre3→ Cre12	2/6*100%	33%
Cre4→ Cre11	2/4*100%	50%
Cre4→ Cre12	2/4*100%	50%
Cre6→ Cre8	2/4*100%	50%
Cre6→ Cre9	2/4*100%	50%
Cre9→ Cre10	3/4*100%	75%
Cre11→ Cre12	3/6*100%	50%

The next process is to form association rules based on table 6. In the case of a strategy to increase sales turnover through promotions by utilizing APRIORI data mining techniques on sales transactions, the minimum confidence value is 50%. Then the form of association rules is obtained as follows as shown in Table 7:

Table 7 Association Rule Table

Combination Pattern 2 items	***Support***	***Confidence***	***Support x Confidence***
Cre1→ Cre2	29%	66%	19,14%
Cre1→ Cre11	21%	50%	10,5%
Cre1→ Cre12	29%	66%	19,14%
Cre2→ Cre3	21%	60%	12,6%
Cre2→ Cre12	21%	60%	12,6%
Cre3→ Cre11	21%	50%	10,5%
Cre4→ Cre11	14%	50%	7%
Cre4→ Cre12	14%	50%	7%
Cre6→ Cre8	14%	50%	7%
Cre6→ Cre9	14%	50%	7%
Cre9→ Cre10	21%	75%	15,8%
Cre11→ Cre12	21%	50%	10,5%

When the three itemsets cannot be recombined, the support calculation process and itemset formation are halted. Association rules are formed from combinations that meet the minimum confidence. Table 7 displays the association rules resulting from the combination of two itemsets. The two best rules that have value confidence is combined with 2 items (Cre1→ Cre2), (Cre1→ Cre12) and (Cre9→ Cre10).

1) If you buy a wardrobe (Cre1), you are likely to buy a dressing table (Cre2) with a support value of 29% and a confidence of 66%.
2) If you buy a wardrobe (Cre1), you will likely buy a springbead (Cre12) with a support value of 29% and a confidence value of 66%. And so on for every combination of 2 items selected in the association rule.
3) If you buy a refrigerator (Cre9), you will likely buy a Kitchen set (Cre10) with a support value of 21% and a confidence of 75%.
4) From the calculation results above, the combination of 2 items with the highest support x confidence value will be used as a combination to determine product promotion at Lido Gallery and for the lowest support x confidence value, it has a small chance to be used as a combination in determining Lido Gallery product promotions. In each product package that is promoted consists of 2 products.

5) In addition to determining the product packages to be promoted based on the calculation results, management can use it to determine the amount of discount on each promotional package.

Conclusion

Based on the study's findings, The a priori algorithm, which has been tested on the strategy of increasing sales turnover through promotion, can be applied, with the sales transaction dataset serving as the source, and the results of the analysis serving as the foundation for making company decisions in the marketing area. The formed association rules can be used as a reference for product recommendations that meet the minimum levels of confidence and support in conducting promotions.

References

C. L. Nangoy, W. J. F. A. Tumbuan, and M. Program, "the Effect of Advertising and Sales Promotion on Consumer Buying Decision of Indovision Tv Cable Provider," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 6, no. 3, pp. 1228–1237, 2018.

S. K. Sinha and P. Verma, "Impact of sales promotion's benefits on brand equity: An empirical investigation," *Glob. Bus. Rev.*, vol. 19, no. 6, pp. 1663–1680, 2018.

D. Hardiyanto, S. Kristiyana, D. A. Sartika, I. Ahmad, and I. Al_Barazanchi, "Training of dual band HT-based emergency communication tools on member of KTB BPBD, yogyakarta," *J. Pengabdi. dan Pemberdaya. Masy. Indones.*, vol. 1, no. 2, pp. 39–47, Oct. 2021.

H. Katole, "Effects of sales promotion campaign adopted by retailers in India," *J. Crit. Rev.*, vol. 7, no. 2, pp. 583–586, 2020.

S. K. Sinha and P. Verma, "The Link Between Sales Promotion's Benefits and Consumers Perception: A Comparative Study Between Rural and Urban Consumers," *Glob. Bus. Rev.*, vol. 20, no. 2, pp. 498–514, 2019.

I. Djamaludin and A. Nursikuwagus, "Analysis of Consumer Purchase Patterns in Sales Transactions Using the Apriori Algorithm," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 8, no. 2, p. 671, 2017.

R. D. Jayapana and Y. Rahayu, "Analysis of consumer buying patterns with a priori algorithm at Rahayu Jepara pharmacy," *UG J.*, vol. Vol. 8, pp. 1–6, 2015.

Z. R. S. Elsi *et al.*, "Utilization of Data Mining Techniques in National Food Security during the Covid-19 Pandemic in Indonesia," *J. Phys. Conf. Ser.*, vol. 1594, no. 1, 2020.

Y. Irawan, R. Wahyuni, Yulisman, and M. B. Punt, "Workshop creating references on the final project of stikes hang tuah pekanbaru students using mendeley application," *J. Pengabdi. dan Pemberdaya. Masy. Indones.*, vol. 1, no. 1, pp. 9–14, Aug. 2021.

S. S. Weng and M. J. Liu, "Feature-based recommendations for one-to-one marketing," *Expert Syst. Appl.*, vol. 26, no. 4, pp. 493–508, 2004.

L. Henando, "Apriori Algorithm and Fp-Growth for Comparative Analysis of Laptop Sales Data Based on Brands Interested by Consumers (Case Study: Indocomputer Payakumbuh)," *J-Click*, vol. 6, no. 2, pp. 201–207, 2019.

L. Hanguang and N. Yu, "Intrusion Detection Technology Research Based on Apriori Algorithm," *Phys. Procedia*, vol. 24, pp. 1615–1620, 2012.

R. Ruswati, A. I. Gufroni, and R. Rianto, "Associative Analysis Data Mining Pattern Against Traffic Accidents Using Apriori Algorithm," *Sci. J. Informatics*, vol. 5, no. 2, pp. 91–104, 2018.

A. Lewis, M. Zarlis, and Z. Situmorang, "Application of Data Mining Using Task Market Basket Analysis on Goods Sales Transactions at AB Mart with Algorithm," *J. Media Inform. Budidarma*, vol. 5, no. April, pp. 676–681, 2021.

M. Bhargava and A. Selwal, "Association Rule mining using Apriori Algorithm: A Review," *Int. J. Adv. Res. Comput. Sci.*, vol. 4, no. 2, p. 2013, 2013.

M. Mittal, S. Pareek, and R. Agarwal, "Loss profit estimation using association rule mining with clustering," *Manag. Sci. Lett.*, vol. 5, no. 2, pp. 167–174, 2015.

M. Hasan, "Journal of Economix Volume 5 Number 2 December 2017," vol. 5, pp. 70–81, 2017.