KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
Vol. 21 (1)(2024), 15-20 p-ISSN: 1693-7554, e-ISSN:2654-3990

Apriori Algorithm Application for Consumer Purchase Patterns Analysis

Boldson Herdianto Situmorang[1,*], **Ali Isra**[2], **Dhatu Paragya**[3], **David Aulia Akbar Adhieputra**[4], [2]

[1]Department of Computer Science, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, West Java, 16143, Indonesia
[2,3,4]Doctoral Programme of Information Technology, Gunadarma University, Depok, West Java, 16424, Indonesia

Abstract

The Apriori algorithm is a data mining association rule algorithm for finding relationship patterns between one or more items in a dataset. Apriori algorithm is often used in transaction data analysis or market basket analysis. Apriori algorithm is used to find out consumer purchase patterns in e-commerce systems and provide product recommendations to consumer by extaracting associations or events from transactional data. This study is purposed for deeply analyze the steps, performance of Apriori algorithm, and give relevant an example of case study to better explain the steps of Apriori algorithm application, as well as the results achieved.

Keywords: *Data mining; association rule; Apriori algorithm, consumer purchase patterns*

1. Introduction

The e-commerce system has significantly changed the conventional commerce system due to developments in information technology and user needs for convenience, fast service, and efficiency in transactions. People who did not use the internet before have to start working online.

To face competitive business competition, an appropriate marketing strategy is needed to attract consumer purchase interest [1]. Data mining is a technique to help entrepreneurs mine information through sales transaction data so that consumer purchase patterns can be identified with the aim of providing product promotion recommendations to consumers. One of the data mining techniques is association, which is a technique for determining frequents itemsets which function to find patterns in data.

The Apriori algorithm is a algorithm of data mining association rule for finding relationship patterns between one or more items in dataset [2]. The Apriori algorithm is often used in transaction data analysis or market basket analysis [3]. By using the Apriori algorithm, entrepreneurs can find out consumer purchase patterns in e-commerce systems and can also provide product recommendations to consumers by extracting associations or events from transactional data [4].

This study is purposed for deeply analyze the steps, advantages and disadvantages of Apriori algorithm, and give relevant an example of case study to better explain the steps of Apriori algorithm application, as well as the results achieved.

[*]**Corresponding author:** ***E-mail adress:*** *boldson.situmorang@unpak.ac.id*
Received: 25 Nov 2023, Accepted: 13 Jan 2024 and available online 19 Jan 2024
https://doi.org/10.33751/komputasi.v21i1.9260

2. Methods

2.1. Data Mining

Data mining is the process of information retrieval by mining data contained in large database to find relationship patterns based on a complex set of data so that it can provide detailed and valuable information [5]. The data analysis stages by implementing data mining can use the Knowledge Discovery in Database (KDD) method which is shown in Figure 1 [6].

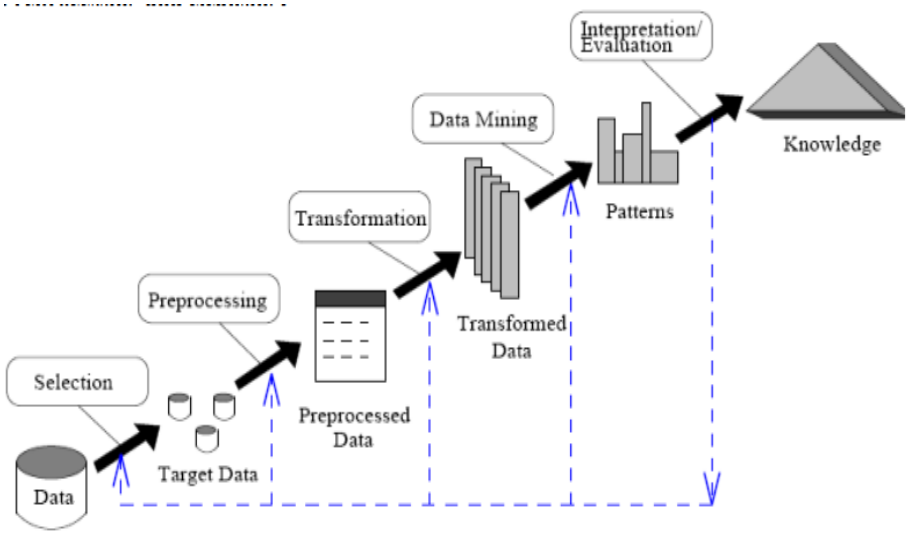

Figure 1. Knowledge Discovery in Database (KDD)

2.2. Apriori Algorithm

The Apriori algorithm is a data mining algorithm with association rules to determine the associative relationship of a combination of items. The relationship between the Apriori algorithm and association is that it can find two or more attributes and two or more objects [7]. The following are data mining association rules for finding high frequency patterns between itemsets [8]:

1. Analyze the high frequency patterns to find combinations of items that meet the minimum support value in database. The support value of an item is obtained using the formula:

$$Support(item) = \frac{\text{The number of transaction contains item}}{\text{Total transaction}} \tag{1}$$

The support value of the 2 items is obtained using the formula:

$$Support(A, B) = P(A \cap B) = \frac{\sum \text{Transactions contains A and B}}{\sum \text{Transactions}} \tag{2}$$

2. Establishment of association rules. After all high frequency patterns are found, then look for association rules that meet the minimum requirements for confidence by calculating the vonfidence value of the associative rules A→B. The confidence value of rule A→B is obtained from the following formula:

$$Confidence(A \rightarrow B) = P(A \mid B) = \frac{\sum \text{Transactions contains A and B}}{\sum \text{Transactions contains A}} \tag{3}$$

Main processes in the Apriori algorithm [9]:

1. Join: In this process, each item is combined with other items until no more combinations are formed.

2. Prune: In this process, the results of the items that have been combined are then pruned using the minimum support determined by the user.

Following are the steps of Apriori algorithm [10]:

1. Apriori uses an iterative approach where k-itemset used to explore (k+1)-itemsets.
2. Candidate (k+1)-itemsets with less frequency or below the threshold (minimum support) will be pruned and not used in determining association rules.
3. The first step is 1-itemset found with perform database scanning to accumulate the number of each item category and its occurrence in every transaction.
4. Next, 1 itemset is used to find 2-itemsets. Potential 2-itemsets candidates are found by pairing one item with another item so that possible combination for 2 items is obtained.
5. Then the 2-itemsets is calculated for the value of its occurance in every transaction. The threshold (minimum support) is determined to pruned out candidates with less frequency.
6. Calculate the support and confidance values for the 2-itemsets that meets the threshold value. 2-itemsets that satisfy minimum support and minimum confidence will be used as association rules.
7. 2-itemsets are used to find 3-itemsets and so on until there are no more frequent (k+1)-itemsets that can be found.
8. After all the rules of the frequent (k+1)-itemsets are formed, then the support and confidence values are calculated.

3. Result and Discussion

3.1. Example of Apriori Algorithm Application

Table 1 shows the purchase records of some products extracted from the transaction database of a cosmetic retail store in Bogor City, West Java Province, Indonesia.

Table 1. Consumer Purchases Record

TID	Consumer Purchases
T1	Highlighter, Foundation, Concealer, Powder, Eyelash Curler, Makeup Setting Spray
T2	Flawless Glow, Highlighter, Blush on, Eyeliner, Lipstick, Eyeshadow, Brushes, Foundation, Face Mask
T3	Concealer, Glitter Eyeshadow, Lip Gloss, Foundation, Highlighter, Eyeshadow
T4	Mascara, Eyeliner, Face Mask, Matte Lip Clay
T5	Lipstick, Concealer, Foundation, Cleansing Balm
T6	Eyeliner, Lip Mask, Foundation, Sponge, Eyeshadow, Concealer
T7	Cleansing Balm, Concealer, Eyebrow, Blush on, Matte Lip Tint
T8	Concealer, Eyebrow, Mascara, Lip Mask, Eyeshadow
T9	Blush on, Mascara, Lip Cream, Cleansing Balm, Lipstick
T10	Sponge, Sunscreen, Concealer, Brushes, Blush on, Highlighter
T11	Foundation, Blush on, Mascara, Sponge, Concealer, Face Mask
T12	Eyeshadow, Lip Cream, Lipstick, Powder, Eyeliner, Foundation, Sunscreen
T13	Lipstick, Cotton Pads Makeup, Powder, Blush on
T14	Brushes, Blush on, Mascara, Lip Gloss, Concealer, Lip Oil, Eye Mask
T15	Foundation, Concealer, Makeup Remover, Face Mask
T16	Brushes, Eyeliner, Blush on, Eye Mask, Eyeshadow, Concealer, Lipstick
T17	Eyebrow, Eyeliner, Cleansing Balm, Concealer, Lipstick
T18	Blush on, Foundation, Mascara, Lipstick, Powder, Highlighter
T19	Highlighter, Blush on, Concealer, Lipstick
T20	Concealer, Eyeliner, Blush on, Eyeshadow, Highlighter

The transaction data is then described in tabular form into 1-itemset as shown in Table 2 with the aim of getting the next (k+1)-itemsets candidates.

Table 2. 1-itemset

No	Itemsets	Quantity	Support
1	Matte Lip Tint	1	5%
2	Eye Mask	2	10%
3	Glitter Eyeshadow	1	5%
4	Makeup Remover	1	5%
5	Setting Spray	1	5%
6	Blush on	11	55%
7	Highlighter	7	35%
8	Sponge	3	15%
9	Lip Mask	2	10%
10	Lip Oil	1	5%
11	Eyebrow	3	15%
12	Eyeshadow	7	35%
13	Lipstick	9	45%
14	Mascara	6	30%
15	Lip Cream	2	10%
16	Eyelash Curler	1	5%
17	Concealer	14	70%
18	Powder	4	20%
19	Brushes	4	20%
20	Cotton Pads Makeup	1	5%
21	Face Mask	4	20%
22	Matte Lip Clay	1	5%
23	Flawless Glow	1	5%
24	Cleansing Balm	4	20%
25	Lip Gloss	2	10%
26	Foundation	9	45%
27	Sunscreen	2	10%
28	Eyeliner	7	35%

No	Itemsets	Quantity	Support
27	Eyeliner	7	35%

For example, in this research the analyst determines minimum support = 30%, then select itemset that has support ≥ minimum support, so the selected data is presented in Table 3.

Table 3. Results of 1-itemset selection

No	Itemsets	Quantity	Support
1	Blush on	11	55%
2	Highlighter	7	35%
3	Eyeshadow	7	35%
4	Lipstick	9	45%
5	Mascara	6	30%
6	Concealer	14	70%
7	Foundation	9	45%
8	Eyeliner	7	35%

Next, select k = k + 1 to determine the 2-itemsets. Then select itemset that meets the minimum support, so the selected data was presented in Table 4.

Table 4. Combination of 2-itemsets

No	Itemsets	Quantity	Support
1	Blush on & Lipstick	6	30%
2	Blush on & Concealer	6	30%
3	Concealer & Foundation	6	30%

Combination of 3-itemsets is processed based on data provided in Table 4, then select itemset that meets the minimum support, so the selected data was presented in Table 5.

Table 5. Combination of 3-itemsets

No	Itemsets	Quantity	Support
1	Blush on Lipstick Concealer	2	10%

The itemset based on Table 5 is not meet the minimum support so that only combination of 2-itemsets is meet the association rules.

Table 6. 2-itemsets selection data

No	Itemsets	Quantity	Confident
1	Blush on Lipstick	30%	54,55%
2	Blush on Concealer	30%	54,55%
3	Concealer Foundation	30%	42,86%

The application of the Apriori Algorithm to the consumer purchases dataset with a minimum support of 30% and a minimum confidence of 50% is presented in Table 7.

Table 7. The results of Apriori algorithm application

Rule	Support	Confident
If buying Blush on then it will buy Lipstick	30%	54,55%
If buying Blush on then it will buy Concealer	30%	54,55%

Based on the association rules resulting from the application of the Apriori algorithm, consumer purchasing patterns can be identified so that product promotion recommendations can be provided to attract customer's buying interest.

Apriori algorithm is not optimized for mining targeted itemsets for class association rule generation in imbalanced data. Apriori algorithm have three main drawbacks which may degrade their performance: (1) the number of candidate itemsets produced during each iteration may be huge and keeping them in memory may be prohibitive; (2) the database needs to be scanned at each iteration, where the number of iterations is equal to the length of the longest frequent itemset; and (3) these algorithms can waste time generating candidate itemsets that do not exist in the database. To address the drawbacks of Apriori algorithm, the Guided Frequent Pattern-Growth (GFP-Growth) algorithm has been proposed to quickly mine a given set of itemsets using a small amount of memory. The problem of mining minority class rules on imbalanced data can be overcome by applying the Minority-Report Algorithm using GFP-Growth [11].

4. Conclusion

The application of the Apriori algorithm through mining association rules for large amounts of data can understand customer consumption habits, analyze the market more effectively, predict sales, control stock of goods, select target customers so that companies can determine the right marketing strategy to attract consumer purchasing power.

Apriori algorithm is not optimized for mining targeted itemsets for class association rule generation in imbalanced data. Apriori algorithm have three main drawbacks which may degrade their performance: (1) the number of candidate itemsets produced during each iteration may be huge and keeping them in memory may be prohibitive; (2) the database needs to be scanned at each iteration, where the number of iterations is equal to the length of the longest frequent itemset; and (3) these algorithms can waste time generating candidate itemsets that do not exist in the database. To address the drawbacks of Apriori algorithm, the Guided Frequent Pattern-Growth (GFP-Growth) algorithm has been proposed to quickly mine a given set of itemsets using a small amount of memory. The problem of mining minority class rules on imbalanced data can be overcome by applying the Minority-Report Algorithm using GFP-Growth.

References

[1] R. Wulandari and I. Mursidah, "Pola Pembelian Produk Berdasarkan Association Rule Data Mining," Semin. Nas. Teknol. Inf. dan Komun. STI&K, vol. 3, no. 1, pp. 265–270, 2019.

[2] S. Siswanti, R. T. Vulandari, A. Kusumaningrum, and S. Setiyowati, "Application of the Apriori Algorithm To Determine the Combination of Poverty Indicators," J. Pilar Nusa Mandiri, vol. 19, no. 1, pp. 45–52, 2023, doi: 10.33480/pilar.v19i1.4161.

[3] M. Qisman, R. Rosadi, and A. S. Abdullah, "Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores)," J. Phys. Conf. Ser., vol. 1722, no. 1, 2021, doi: 10.1088/1742-6596/1722/1/012020.

[4] A. D. Hartanto, B. C. Lim, and D. Pradana, "Apriori Algorthm Implementation to Determine Product Sales Priority," CCIT (Creative Commun. Innov. Technol. J., vol. 13, no. 1, pp. 1–9, 2020.

[5] A. Rahmadsyah and N. Mayasari, "Grouping Goods with the Association Rule Method Using A priori Algorithms in Modern Retail Stores," Int. J. Sci. Technol. Manag., vol. 3, no. 6, pp. 1527–1532, 2022, doi: 10.46729/ijstm.v3i6.637.

[6] C. Zhang and J. Han, "Data Mining and Knowledge Discovery," Urban B. Ser., vol. 6, no. 1, pp. 797–814, 2021, doi: 10.1007/978-981-15-8983-6-42.

[7] P. H. Simbolon, "Implementasi Data Mining Pada Sistem Persediaan Barang Menggunakan Algoritma Apriori (Studi Kasus: Srikandi Cash Credit Elektronic dan Furniture)," Jurikom), vol. 6, no. 4, pp. 401–406, 2019, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C401.

[8] B. I. S. P. Putu, N. P. S. M. Suryan, and A. Sri, "Analysis of Apriori Algorithm on Sales Transactions to Arrange Placement of Goods on Minimarket," IJEET Int. J. Eng. Emerg. Technol., vol. 3, no. 1, pp. 13–17, 2018.

[9] C. S. Fatoni, E. Utami, and F. W. Wibowo, "Online Store Product Recommendation System Uses Apriori Method," J. Phys. Conf. Ser., vol. 1140, no. 1, 2018, doi: 10.1088/1742-6596/1140/1/012034

[10] E. Irfiani, "Application of Apriori Algorithms to Determine Associations in Outdoor Sports Equipment Stores," SinkrOn, vol. 3, no. 2, p. 218, 2019, doi: 10.33395/sinkron.v3i2.10089.

[11] L. Shabtay, P. Fournier-Viger, R. Yaari, and I. Dattner, "A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data," Inf. Sci. (Ny)., vol. 553, pp. 353–375, 2021, doi: 10.1016/j.ins.2020.10.020