Journal of Computer Science and Information Technology
https://jcsitech-upiyptk.org/ojs
2024 Volume 10 Issue 2 Pages 39-43 e-ISSN: 2502-1486

Shopping Cart Analysis to Support Business Management with the Apriori Algorithm

Nurul Aini*, Eka Praja Wiyata Mandala, Muhammad Afdhal
Universitas Putra Indonesia YPTK Padang, Jl. Raya Lubuk Begalung Padang, Sumatera Barat – 25221, Indonesia
* nurulaini120400@gmail.com

Abstract

The increasingly advanced development of the business world has led to increasingly fierce competition. One way to maintain the company's survival is to maintain good relationships with customers. Each market has its own way of increasing sales. Serambi Mart is one of the minimarkets operating in Batusangkar City. This minimarket is relatively new because it has been operating since 2020. Even though it is new, this minimarket is quite busy with consumers visiting for shopping, this is due to Serambi Mart management choosing the right location. Based on observations made, product layout arrangements are still based on subjective management, so there are several products that are not suitable to be compared. The layout seems messy, causing difficulties for consumers in shopping. By utilizing sales transaction data at Serambi Mart. This research uses the Apriori algorithm. The Apriori algorithm is an association rule and looks for relationship patterns between one or more items in data, using Association Rules with Minimum Support of 30% and Minimum Confidence of 50%. Therefore, an application is needed that can help Serambi Mart to get information. One way to get this information is to utilize data mining techniques. By using the Apriori Algorithm method for arranging the product layout at Serambi Mart, it is hoped that it can provide convenience to consumers who shop.

Keywords: Business Management, Data Mining, Apriori Algorithm, Association Rules, Sales

1. Introduction

In previous research conducted by Kristoko Dwi Hartomo, Sri Yulianto, and Rahmat Abadi Suharjo in 2020 with the title Stock Prediction and Arrangement of Goods Layout Using a Combination of the Triple Exponential Smoothing and FP-Growth Algorithms, it was explained that the data processing in this research used the Triple Algorithm. Exponential Smoothing and FP-Growth. The Triple Exponential Smoothing algorithm is a prediction algorithm for managing stock items. The FP-Growth algorithm is used to determine consumer behavior patterns so that it is used for decision making in preparing goods. The results of research using the FP-Growth algorithm found 12 association rules, the association rules that had the highest lift ratio values were tea and sugar with a value of 6,131. With the Triple Exponential Smoothing algorithm, the prediction results for 2018 were 131,141 kg with a MAPE accuracy level of 88.3 % [1].

In another research conducted by Alfiqra, Faiza Yogi Alfizi in 2018 with the title Application of Market Basket Analysis Using the KDD (Knowledge Discovery In Databases) Process as a Supermarket Product Sales Strategy (Case Study: Supermarket X), this research explains that this research uses the FP-Growth method to obtain association rule results from a set of data. The KDD process in this research explains systematically the search for new relationships in market basket analysis using several stages of data processing. The steps in the KDD process include selection, preprocessing, transformation, data mining, and interpretation/evaluation. This research succeeded in obtaining 10 association rules that explain consumer shopping habits, with a support value of 12% and a confidence value of 66.7 % , namely snack and beverage department products [2]. By utilizing sales transaction data at Serambi Mart and processing it into very useful and profitable information. Therefore, an application is needed that can help Serambi Mart to get information. One way to get this information is to utilize data mining techniques.

Then the analysis stage was carried out using data mining techniques using the a priori algorithm method to obtain information. The a priori algorithm is a data retrieval algorithm with association rules to determine the associative relationship of a combination of items such as which products are purchased by consumers simultaneously, using support and confidence data mining techniques, where the support value shows how big the level of dominance of an item is in the overall transaction and The confidence value shows the relationship between two items based on certain conditions [3] [4].

Management in general is the art and science of planning, organizing, directing and supervising the efforts of an organization and the use of organizational

Accepted: 02-05-2024 | Revised: 03-10-2024 | Published: 04-30-2024 | doi: 10.35134/jcsitech.v10i2.98

resources that can be used to achieve predetermined goal. The definition of management according to experts is a science that studies understanding why and how humans work to produce benefits for themselves [5].

Business is a business activity to achieve the needs of society, humans and an organization. An entrepreneur must be astute in seeing needs and changes or shifts in the needs of society itself. The emergence of a new need or the emergence of changes to existing needs is a business opportunity that can be utilized properly to generate profits [6].

Information technology is a technology that is used for data processing, data processing and data storage in various ways and execution to produce quality information [7][8]. Quality information in question is relevant, timely, accurate information that can be used by individuals, business organizations or governments and is strategic information and can be used as a tool in making decisions [9] [10].

Data mining is a process for producing useful data information from large databases. Data mining is also called Knowledge Discovery in Database or abbreviated as KKD. The KDD process links the results of the process of extracting trends in data patterns, then the results are carefully converted into information that is easy to understand [11].

Data mining is a semi-automatic system that uses artificial intelligence, mathematics, statistics and machine learning techniques to identify and extract useful knowledge information and is stored in large databases [12][13].

One method used to look for patterns that often appear among many transactions, where each transaction consists of several items, is the definition of Association rule. This association rule produces how big the relationship is between X and Y, and two measures are needed for this rule, namely support and confidence. Support is the percentage of combinations of an item in the database or supporting value. Confidence is a certainty value that can be sought after the frequency pattern of the appearance of an item is found [14][15].

A priori algorithms are used to obtain association rules and look for relationship patterns between one or more items in data. One of the stages in association analysis that has attracted the attention of many researchers is to produce an efficient algorithm, namely frequency mining patterns [16]. Whether an association is important or not can be measured using a comparison of support and confidence. The support value is a value that shows the level of dominance of an item in the entire transaction, while the confidence or certainty value is the strength of the relationship between items in the association rules [17].

2. Research methodology

The research framework is the concept or stages that will be carried out in the research. So that the steps taken by the author in this design do not deviate from the main discussion and are easier to understand, the sequence of research steps will be made systematically so that it can be used as a clear and easy guide for solving existing problems. The research framework that the author carried out in the research can be depicted in Figure 1 .

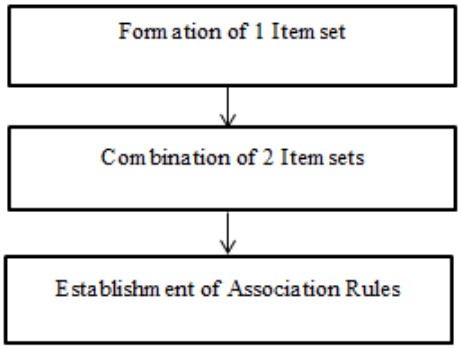

Figure 1. Research Framework

This stage looks for a combination of items that meet the minimum requirements for the support value of the database. The support value of an item is obtained using the following formula:

$$Support\ (A) = \frac{\sum Jumlah\ Transaksi\ Mengandung\ A}{\sum Total\ Transaksi} \times 100\%$$

The formula above is to search for itemset 1, while to search for itemset 2, use the following formula.

$$Support\ (A \cap B) = \frac{\sum Jumlah\ Transaksi\ Mengandung\ A\ dan\ B}{\sum Total\ Transaksi} \times 100\%$$

The formula above is to search for itemset 2, while to search for itemset 3 use the following formula

$$Support\ (A,B,C) = P(A \cap B \cap C)$$

$$Support\ (A,B,C) = \frac{\sum transaksi\ mengandung\ A,B\ dan\ C}{\sum \text{transaksi}} \times 100\%$$

After all the frequency patterns are found, then look for the associative rule that meets the minimum requirements for confidence by calculating *the confidence of the associative* rule A → B obtained from the following formula:

$$Confidence\ P\left(\frac{B}{A}\right) = \frac{\sum Hasil\ Support\ Mengandung\ A\ dan\ B}{\sum Total\ Support} \times 100\%$$

3. Results and Discussion

3.1 Data analysis

The IT Business Management system that will be designed at Serambi Mart can store sales transaction data. From this data, data analysis and processing will be carried out using one of the data mining calculations, namely a priori.

Below are 20 transaction samples taken randomly at Serambi Mart which can be seen in the following table:

Table 1 . Product Sales Transactions

No	Items purchased
1	Big blue ABC battery, fried Indomie, Aqua 600 ml, Beng-beng 22g
2	Buavita orange, Choki-choki, Ademsari chinku 350 ml bottle, Indomie fried
3	Beng-beng 22g, Bear brand, Aqua 600 ml, Sampoerna cigarettes 16
4	Fried Indomie, Coca Cola 250 ml, Sampoerna 16 cigarettes, Aqua 600 ml
5	1 kg granulated sugar, fried Indomie, 22g Beng-beng, 600 ml Aqua
6	Aqua 600 ml, LPG Gas 3 kg, Beng-beng 22g, Sampoerna cigarettes 16
7	Sampoerna cigarettes 16, Granulated sugar 1 kg, Sosro bottle 450 ml, Aqua 600 ml
8	Beng-beng 22g, Aqua 600 ml, fried Indomie, granulated sugar 1 kg
9	KF94 mask, fried Indomie, Sampoerna cigarettes 16, Aqua 600 ml
10	Aqua 600 ml, Sari white bread, Sugar 1 kg, Sampoerna cigarettes 16
11	Beng-beng 22g, Frisian flag uht low fat vanilla, Pepsodent Action 123 siwak 150g, fried Indomie
12	Frisian flag uht low fat vanilla, Aqua 600 ml, Kit-kat mini, Granulated sugar 1 kg
13	Aqua 600 ml, Frisian flag uht low fat vanilla, LPG gas 3 kg, Sampoerna cigarettes 16
14	SGMexplor 1+vanilla 400g, Kit-kat mini, fried Indomie, Sunlight lime 210 ml
15	Rexona shower clean 50ml, Daia white 280g, Blue triangle flour 1 kg, Granulated sugar 1 kg
16	Tango long chocolate 130g, Beng-beng 22g, Bebelove 400g, Aqua 600 ml
17	Coca Cola 250 ml, Indomie fried, Beng-beng 22g, Sampoerna cigarettes 16
18	Sampoerna 16 cigarettes, LPG gas 3 kg, Beng-beng 22g, Good day original cappuccino
19	Sari roti plain, Bintang zero 330 ml, Frisian flag uht low fat vanilla, Aqua 600 ml
20	Tango long chocolate 130g, Aqua 600 ml, Granulated sugar 1 kg, Eucalyptus oil 210 ml

Based on the data above, the data can be processed and then applied in the a priori algorithm stages, obtaining data with the following calculation results:

a. Itemset formation

The process of forming C1 or what is called 1 itemset with a minimum amount of support = 30%. The following is the calculation for forming 1 itemset:

Support (Battery)= (1)/20×100%=5%
Support (Aqua)= (14)/20×100%=70%
Support (Bebelove)= (1)/20×100%=5%
Support (Buavita)= (1)/20×100%=5%
Support (Choki-choki)= (1)/20×100%=5%
Support (Beng-beng)= (9)/20×100%=45%
Support (Bear brand)= 1/20×100%=45%
Support (Cigarettes)= (9)/20×100%=45%
Support (Instant noodles)= (9)/20×100%=45%
Support (Coca cola)= (2)/20×100%=10%
Support (Granulated sugar)= (7)/20×100%=35%
Support (Ademsari)= (1)/20×100%=5%
Support (LPG Gas)= (3)/20×100%=15%
Support (Mask)= (1)/20×100%=5%
Support (Bread essence)= (2)/20×100%=20%
Support (Sunlight)= (1)/20×100%=5%
Support (Milk box)= (4)/20×100%=20%
Support (Tea bottle)= (1)/20×100%=5%
Support (Kit-kat)= (2)/20×100%=10%
Support (Pepsodent)= (1)/20×100%=5%
Support (SGM)= (1)/20×100%=5%
Support (Rexona)= (1)/20×100%=5%
Support (Daia)= (1)/20×100%=5%
Support (Tango)= (2)/20×100%=10%
Support (Flour)= (1)/20×100%=5%
Support (Good day)= (1)/20×100%=5%
Support (Zero star)= (1)/20×100%=5%
Support (Eucalyptus oil)= (1)/20×100%=5%

The minimum support specified is 30%, itemsets that do not meet the minimum support are removed, as in the table below :

Table 2. Minimum Support

No	Product	Probability	*Support*		*Percentage*
1	Aqua	14	14/20	0.7	70%
2	Beng-beng	9	9/20	0.45	45%
3	Cigarette	9	9/20	0.45	45%
4	Instant noodles	9	9/20	0.45	45%
5	Sugar	7	7/20	0.35	35%

b. Combination of two itemsets

The process of forming C2 or what is called 2 itemsets with a minimum amount of support = 15%. Based on the formula above, the following calculations can be made:

Support (Aqua ,Beng -beng)=(6)/20×100%=30%
Support (Aqua ,Cigarette)=(7)/20×100%=35%
Support (Aqua , instant noodles)=(5)/20×100%=25%
Support (Aqua ,Sugar)=(6)/20×100%=30%
Support (Beng-beng , Cigarettes)=(4)/20×100%=20%
Support (Beng-beng , instant noodles)=(5)/20×100%=25%
Support (Beng-beng , Sugar)=2/20×100%=10%
Support (Cigarettes , instant noodles)=(3)/20×100%=15%
Support (Cigarettes , Sugar)=(2)/20×100%=10%
Support (instant noodles , sugar)=2/20×100%=10%

The minimum support specified is 30%, itemsets that do not meet the minimum support will be removed, as in the table below :

Table 3. *Frequent Itemset* 2

No	Product	Probability	*Support*		*Percentage*
1	Aqua, bro	6	6/20	0.3	30%
2	Aqua, Cigarettes	7	7/20	0.35	35%
3	Aqua, Sugar	6	6/20	0.3	30%

c. Combination of Three Itemsets

The process of forming C3 or what is called 3 itemsets with a minimum amount of *support* = 30%. Following are the calculations for 3 itemsets:

$$Support\ (Aqua, Beng - beng, Rokok) = \frac{2}{20} \times 100\% = 10\%$$

$$Support\ (Aqua, Beng - beng, Gula) = \frac{2}{20} \times 100\% = 10\%$$

$$Support\ (Aqua, Rokok, Gula) = \frac{2}{20} \times 100\% = 10\%$$

The minimum support specified is 30%, in the search 3 itemsets did not meet the minimum support. So, to form an itemset, stop combining 2 itemsets.

After all frequency patterns high is found, followed by looking for association rules that meet *the minimum requirements confidence* by calculating *the confidence of* the associative rule $A \rightarrow B$. Minimum *confidence* = 50%. The *confidence* value of the association rule $A \rightarrow B$ is seen calculation to find the *confidence* value :

$$Confidence = \mathrm{P} = \frac{\sum (\mathrm{Aqua, Beng - beng})\_0{,}3}{\sum \mathrm{Aqua}\qquad 0{,}7} \times 100\% = 0{,}4285 = 42{,}85\%$$

$$Confidence = \mathrm{P} = \frac{\sum (\mathrm{Aqua, Beng - beng})\_0{,}3}{\sum \mathrm{Beng - beng}\qquad 0{,}45} \times 100\% = 0{,}6667 = 66{,}67\%$$

$$Confidence = \mathrm{P} = \frac{\sum (\mathrm{Aqua, Rokok})\_0{,}35}{\sum Aqua\qquad 0{,}7} \times 100\% = 0{,}5 = 50\%$$

$$Confidence = \mathrm{P} = \frac{\sum (\mathrm{Aqua, Rokok})\_0{,}35}{\sum Rokok\qquad 0{,}45} \times 100\% = 0{,}7778 = 77{,}78\%$$

$$Confidence = \mathrm{P} = \frac{\sum (\mathrm{Aqua, Gula})\_0{,}3}{\sum Aqua\qquad 0{,}7} \times 100\% = 0{,}4285 = 42{,}85\%$$

$$Confidence = \mathrm{P} = \frac{\sum (\mathrm{Aqua, Gula})\_0{,}3}{\sum Gula\qquad 0{,}35} \times 100\% = 0{,}8571 = 85{,}71\%$$

Search results for *confidence* obtained from calculating a combination of 2 *itemsets* with a minimum confidence value of 50%, then those that do not meet the min *confidence* are not displayed in the table contained in the following table:

Table 4. Association Rules

Knowledge	*Confidence*	*Percentage*	Information
- If a customer buys Beng-beng, then buy Aqua	0.6667	66.67%	Rule – 01
- If a customer buys Aqua, they buy cigarettes	0.5	50%	Rule – 02
- If a customer buys cigarettes, then buy Aqua	0.7778	77.78%	Rule-03
- If the Customer buys Sugar, then buy Aqua	0.8571	85.71%	Rule-04

4. Conclusion

The application of a priori algorithms to the system that has been created is able to recognize customer buying patterns to increase sales. This is proven by the success of the a priori algorithm method in analyzing high frequency patterns with a predetermined minimum support of 30% starting from finding the support value in itemset 1, then itemset 2 and continuing with itemset 3. After obtaining all the high frequency patterns, then proceed by looking for association rules that meet the minimum confidence requirements by calculating the confidence of associative rules. The search results for confidence obtained from calculating a combination of 2 itemsets then obtained results from 4 association rules.

References

[1] Hartomo, KD, Yulianto, S., & Suharjo, RA (2020). Stock Prediction and Item Layout Settings Using a Combination of Triple Exponential Smoothing and FP-Growth Algorithms. Journal of Information Technology and Computer Science (JTIIK), 7(5), 869-878. https://doi.org/10.2516/jtiik.202071863

[2] Alfiqra, A., & Alfizi, FY (2018). Implementation of Market Basket Analysis Using the KDD (Knowledge Discovery in Database) Process as a Self-Service Product Sales Strategy (Case Study: Supermarket X). IENACO (Industrial Engineering National Conference) 6 2018.

[3] Saputra, R., & Sibarani, AJ (2020). Implementation of Data Mining Using the Apriori Algorithm to Improve Drug Sales Patterns. JATISI (Journal of Information Engineering and Information Systems), 7(2), 262-276. https://doi.org/10.35957/jatisi.v7i2.195

[4] Riszky, AR, & Sadikin, M. (2019). Data Mining Using Apriori Algorithms for Product Recommendations for Customers. Journal of Computer Technology and Systems, 7(3), 103-108. https://doi.org/10.14710/jtsiskom.7.3.2019.103-108

[5] Darim, A. (2020). Organizational behavior management in creating competent human resources. Munaddhomah: Journal of Islamic Education Management, 1(1), 22-40. https://doi.org/10.31538/munaddhomah.v1i1.29

[6] Luckandi, D. (2019). Analysis of Payment Transactions Using Fintech in MSMEs in Indonesia: Adaptive Structuration Theory Approach. JIKO (Journal of Informatics and Computers), 4(1), 1-10. http://dx.doi.org/10.26798/jiko.v4i1.126

[7] Zikri, A., Adrian, J., Soniawan, A., Azim, R., Dinur, R., & Akbar, R. (2017). Implementation of Business Intelligence to Analyze Child Birth Data at the Ani Padang Clinic using the Tableau Public Application. Online Journal of Informatics, 2(1), 20-24. https://doi.org/10.15575/join.v2i1.70

[8] Yuliana, K., Saryani, S., & Azizah, N. (2019). Web-Based Goods Delivery Recapitulation Design. Global Sysphotech Journal, 9(1).

[9] Cholik, C. A. (2021). Development of information communication technology/ICT in various fields. Journal of the Faculty of Brass Engineering, 2(2), 39-46.

[10] Naibaho, RS (2017). The role and planning of information technology in the company. Warta Dharmawangsa, (52). https://doi.org/10.46576/wdw.v0i52.253

[11] Mardi, Y. (2017). Data mining: Classification using the c4 algorithm. 5. Edic Journal of Informatics Research in the Field of Computer Science and Informatics Education, 2(2), 213-219. https://doi.org/10.22202/ei.2016.v2i2.1465

[12] Elfaladonna, F., & Rahmadani, A. (2019). Analysis of the Classification-Decision Tree Method and the C. 45 Algorithm to Predict Diabetes Using the Rapid Miner Application. SINTECH (Science and Information Technology) Journal, 2(1), 10-17. https://doi.org/10.31598/sintechjournal.v2i1.293

[13] Pramadani, E., Sunandar, H., & Hasan, Y. (2019). Implementation of Newspaper Sales Data Mining Using the C4 Method. 5 (Case Study: Pt. Media Mass Cahaya Renewal). Scientific Information and Technology (CORE), 6(1), 5-10.

[14] Mulya, DP (2019). Analysis and Implementation of Association Rules with the Fp-Growth Algorithm in Selecting Clay Soil Purchases (Case Study at Pt. Anveve Ismi Berjaya). Journal of Technology and Business Information Systems, 1(1), 47-57. http://dx.doi.org/10.30865/mib.v4i4.2300

[15] Simanjuntak, HE, & Windarto, W. (2020). Data Mining Analysis Using Frequent Pattern Growth on PT Mora Telematics Indonesia Sales Transaction Data for Internet Product Marketing Strategy Recommendations. Budidarma Media Informatics Journal, 4(4), 914-923. https://doi.org/10.47233/jteksis.v1i1.6

[16] Kusumo, H., Sediyono, E., & Marwata, M. (2019). A priori algorithm analysis to support higher education promotion strategies. Walisongo J. Inf. Technol, 1(1), 49. https://doi.org/10.21580/wjit.2019.1.1.4000

[17] Simbolon, PH (2019). Implementation of Data Mining in the Goods Inventory System Using the Apriori Algorithm (Case Study: Srikandi Cash Credit Electronics and Furniture). JURIKOM (Journal of Computer Research), 6(4), 401-406. http://dx.doi.org/10.30865/jurikom.v6i4.1399