

Association Relationship Analysis in Finding Sales of Goods With Apriori Algorithm

Humam Fathurrahman^{1)*}, Aswan S. Sunge²⁾, Sufajar Butsianto³⁾

^{1)*2)3)}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Bekasi, Indonesia

^{1)*}humam.fathur@mhs.pelitabangsa.ac.id, ²⁾aswan.sunge@pelitabangsa.ac.id, ³⁾sufajar.s@pelitabangsa.ac.id

ABSTRACT

Technology can be designed to help human life from all aspects ranging from agriculture, health science, industry and daily life. Toko Intan, a business engaged in the sale of basic and daily necessities. Every day, Toko Intan records every sales transaction in an archive stored in Microsoft Excel, containing data on goods sold every day. The purpose of this research is to find out what items are bought simultaneously by consumers to manage inventory, with the data mining method used in this research is the Association Rules method. Association Rules is one of the data mining techniques from the a priori algorithm which functions to find a combination pattern of an item. Tests carried out to process data in this study using the RapidMiner application, from tests carried out with the specified parameters, namely minimum support 30% and minimum confidence 65%, resulted in a lift ratio validation rule of 1.206. Personal Care Biscuits with 30.8% support and 90.9% confidence with a validation lift ratio of 1.206. Sales transaction data analysis can be applied well, and is able to generate a new association rule from the sales transaction dataset. With this research, it is hoped that it can provide information to the owner of the Intan store in providing the stock of goods needed by consumers and to find out the combination of item sets from the sales transaction dataset.

Keywords: Sales; Apriori; Association rules; Data Mining; RapidMiner

1. INTRODUCTION

Technology is designed to help human life from all aspects ranging from agriculture, health science, industry and daily life, with technology every activity is carried out more easily, quickly and systemized so it is not uncommon for every business world and the industrial world today to compete to continue to develop technological systems to be applied in every activity, because technology can also save in the production cost process.

Intan Store is one of the stores engaged in selling various types of daily necessities. With daily sales activities, the sales data of goods will increase more and more. Data does not only function as an archive for the store, data can be utilized and processed into useful information to increase product sales. Some of the problems that often arise regarding the sale of goods are that it is difficult for stores to identify items that are less desirable or not sold within a certain period of time. The availability of large sales data is not used optimally because there is no decision support system and method that can be used to design a business strategy to increase sales.

Sales are the processes, methods, and behaviors associated with selling. Selling means giving something to a buyer for payment or collection (Kristania & Listanto, 2022) (Faqih & Wahyudi, 2022). Sales are very important and play a vital role in the business world, especially in trade, to maintain and develop the business in order to make a profit and maintain the existence of the business owned. In general, sales can be defined as buying and selling activities involving two or more parties using legal tender (Syabania & Rosmawarni, 2021).

Goods are objects or entities offered to buyers, both tangible and intangible, which consumers can accept to meet needs and provide satisfaction (Khanza et al., 2021). Goods have a tangible form, are touched directly, have a value that can be measured economically usually in the form of a price or exchange rate and goods can be moved from one place to another.

Association rule is a data mining technique to find associative rules or combination patterns of items (A Fadila Shely et al., 2021). Based on this definition, association rules are combinations of items that meet the minimum support value requirements in a database (Lienata et al., 2021). This technique is particularly useful in market basket analysis, where we want to know what products customers often buy together.

Data Mining is the activity of extracting data from a very large data set to find information that has its own uses

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

according to the needs of the data. information that has its own usefulness according to the needs (Annisa Ilham Fathimah et al., 2022). To find insights or a certain pattern in a large amount of data (H Arva Abhyoso et al., 2020). Apriori algorithm is a type of association rule in data mining (Triayudi, 2022). Establishment of item set relationships based on the support value at a predetermined frequency for the process of establishing 2-item relationships, 3-item relationships or other item relationships (Triayudi & Iskandar, 2022). This process involves identifying combinations of items that frequently occur together in a dataset, allowing businesses to understand customer purchasing behavior and optimize product placement, marketing strategies, and inventory management. The Apriori algorithm utilizes knowledge about the frequency of previously known attributes to process subsequent information. In Apriori algorithm, possible candidates are determined by considering the minimum support and minimum confidence values. Support is the percentage of combinations of items in the database (Triyuliani et al., 2020).

The implementation of the Apriori data mining method in sales at Alfamart. The study focuses on determining product patterns and popularity to assist buyers in making informed purchasing decisions. Data analysis was conducted on sales data from Alfamart in January 2020, with a minimum support of 10 and a confidence of 50%. The analysis process involved interviews, documentation, and transcription of sales transactions. The study utilized the Apriori algorithm to find frequent itemsets in transaction data. RapidMiner was used for data mining, text analysis, and pattern extraction. The research highlights the importance of data analysis tools for businesses to optimize product placement and meet consumer needs (Agustiani et al., 2020).

2. LITERATURE REVIEW

The implementation of the Apriori algorithm in data mining to analyze food product sales at Kopsyahira, aiming to identify associated sales patterns and improve stock management. Data mining, as part of Knowledge Discovery in Databases (KDD), involves selecting and processing data to discover meaningful relationships and trends. The methodology includes data collection, transformation, attribute selection, dataset creation, Apriori rule establishment, and analysis using Tanagra software. The study emphasizes the importance of efficient decision-making systems to enhance sales and service quality (Riyanto & Susanti, 2020). Implemented data mining using the Apriori algorithm to analyze home snack sales data at DAPOERIN'S small business. Using sales transaction data during February 2020 and a minimum support of 30%, this study successfully identified the best-selling products (Lontong, Putu Ayu, Risoles, and Piscok) and found patterns of product purchase associations. The main results show that the purchase of Putu Ayu, Risoles, or Piscok tends to be followed by the purchase of Lontong, with a high confidence level. Although limited to a one-month data sample, this study demonstrates the benefits of applying data mining for sales analysis in small businesses, which can help in inventory optimization, marketing strategy development, and better business decision-making (Prasetyo et al., 2020). Applied the Apriori algorithm to analyze sales transaction data at a tent stall (warten) to recommend food and beverage menu combinations. The researchers collected sales transaction data for one day and analyzed it using Apriori algorithm with minimum support 20% and minimum confidence 70%. The analysis resulted in 11 association rules with 2-itemsets and 1 rule with 3-itemsets that co-occurred most frequently. The menu combination with the highest support and confidence values is Es Teh Manis and Mendoan (50% support, 76% confidence). This research aims to help tent shop owners in making menu packages that match customer preferences, so as to increase sales and customer satisfaction. The methods used include data collection, identification of business problems, analysis using the Apriori algorithm, and conclusion drawing. This research shows that the Apriori algorithm can be effectively applied to analyze customer purchasing patterns and provide potential menu combination recommendations for culinary businesses (M Nanda Nurisya et al., 2022).

3. METHOD

Data Mining Processing

Before carrying out the data mining process using the apriori algorithm or association rule, it is necessary to process the data first in order to get the appropriate and desired dataset. Data mining processing in this study follows the stages of Knowledge Discovery in Database (KDD) to produce information in accordance with a predetermined order. The data mining processing in this research follows the stages of KDD to produce useful information. The following are the stages carried out:

* Corresponding author



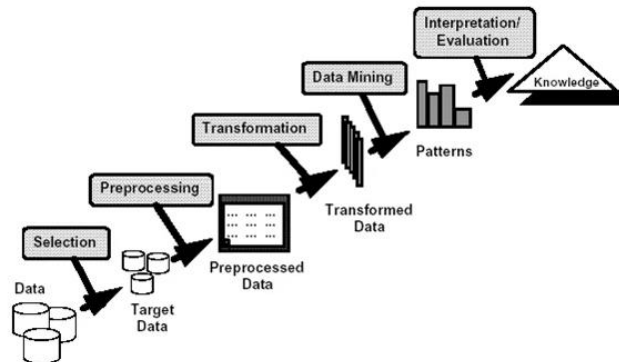


Fig. 1 KDD Steps in Data Mining

1. Data Selection

Data Selection is the first step in processing data from sales samples that will be used for the data mining process.

2. Preprocessing

In the preprocessing stage for this research, several cases were found where the same product appeared more than once in one transaction number. The preprocessing stage involved identifying each case of duplicated products in a particular transaction. The duplicated products were then removed to ensure the integrity and accuracy of the data used in the analysis. This stage is important to avoid inaccurate analysis results due to data duplication, so that each transaction no gets unique and relevant information needed for the next stage.

3. Transformation

Transformation is an important step in the process of extracting valuable information from raw data. Data transformation involves various techniques and methods to transform, clean and format data to make it more suitable for analysis. Transformation stage, the dataset that has been obtained from the previous process is converted into a format that can be used for the data mining process.

4. Data Mining

Data mining process requires a tool that helps execute data mining operations based on the analysis model that has been defined. This data mining processing performs an analysis process on data by emphasizing the discovery of information on very large and hidden types of data stored in a database.

5. Evaluation

Part of the components that interact with data mining modules. It is also part of the software that finds patterns in the database that are processed in data mining analysis so as to find the appropriate knowledge (Saefudin & F Donny, 2020).

Process Of Association Rule Method

Association analysis is defined as the process of finding all association rules that meet the minimum requirements for support and the minimum requirements for confidence. Support is the percentage of such item combinations in the database, while confidence is the strength of the relationship between items in the association rule (Sahara et al., 2022). The support value of an item is obtained using the following formula :

$$\text{Support (A)} = \frac{\sum \text{Transactions containing A}}{\sum \text{Transactions}} \times 100\% \tag{1}$$

The Support value of two items is obtained from the following formula :

* Corresponding author



$$\text{Support (A, B)} = \frac{\sum \text{Transactions containing A and B}}{\sum \text{Transactions}} \times 100\% \quad (2)$$

After all the high-frequency patterns are found, the association rule that meets the minimum requirement for confidence is found by calculating the confidence of the associative rule $A \cap B$. The confidence value of rule $A \cap B$ is obtained by the following formula :

$$\text{Confidence} = P(B|A) = \frac{\sum \text{Transactions containing A and B}}{\sum \text{Transactions containing A}} \times 100\% \quad (3)$$

To determine the association rule to be selected, it must be sorted based on $\text{Support} \times \text{Confidence}$. Rules are taken as many as n rules that have the largest results (Takdirillah, 2020). The Lift Ratio test is a method used to measure the strength of the association that has been formed. Lift Ratio can be used as a determinant of the strength and validity of association patterns. The following is the Lift Ratio formula :

$$\text{Lift Ratio} = \frac{\text{Confidence}}{\text{Benchmark Confidence}} \quad (4)$$

Benchmark Confidence is calculated by the following formula as follows :

$$\text{Benchmark Confidence} = \frac{\text{Transactions containing Consequent}}{\sum \text{Transactions}} \times 100\% \quad (5)$$

4. RESULT

The data used in this research comes from Intan Store, a business engaged in the sale of basic and daily necessities. This data is in the form of daily sales transaction records stored in Microsoft Excel format. Each transaction includes information about the transaction number, product, quantity, category, price, and total. The data collection process is done by downloading data from the sales system. The data used is the original data recorded by Intan Store during a one-month period.

Association Rule Calculation Result

Based on the analysis conducted, it was found that when "Wafer" is the antecedent, "Biscuit" as the consequent results in a lift ratio of 0.874. This lift ratio is less than 1, which indicates the relationship between "Wafer" and "Biscuit" is invalid. In addition, when "Personal Care" is the antecedent, "Biscuits" as the consequent results in a lift ratio of 1.206. This lift ratio is also more than 1, indicating the relationship between "Personal Care" and "Biscuits" is valid.

Table 1
Result of Lift Ratio Calculation

Antecedent	Consequent	Confidence %	Benchmark Confidence %	Lift Ratio
Wafer	Biskuit	65,9	75,4	0,874
Personal Care	Biskuit	90,9	75,4	1,206

Calculation Results With Rapidminer Application

In this process, the a priori algorithm method is applied to find the relationship between item sets with the right level of accuracy. In this study, the authors used RapidMiner as a tool for testing.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convict...
1	Wafer	Biskuit	0.446	0.659	0.862	-0.908	-0.064	0.874	0.722
2	Personal Care	Biskuit	0.308	0.909	0.977	-0.369	0.053	1.206	2.708

Fig. 2 Data Conclusion Process Results with Rapidminer Application

* Corresponding author



Figure 2 above displays a detailed visualization of the combinations of items that frequently co-occur in the analyzed dataset. In addition, it also presents the support, confidence and lift values used to evaluate the strength and significance of the associative rules found.

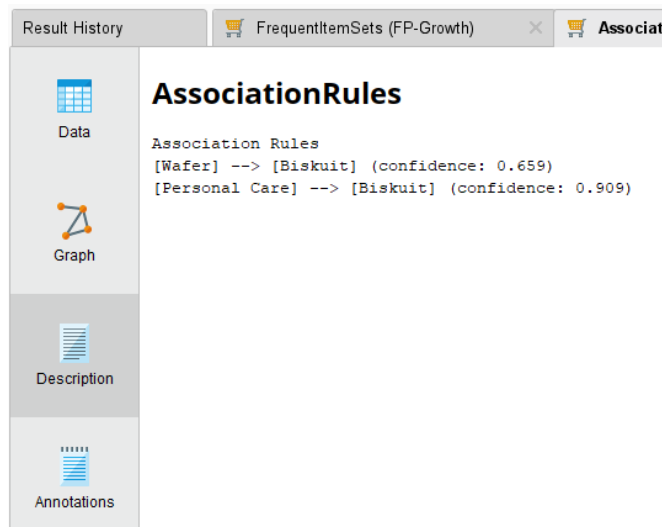


Fig. 3 Association Rules Formed With Rapidminer Application

Figure 3 above shows the significant associative rules found in the transaction data. This information can be used to understand customer buying behavior and optimize marketing strategies and product placement in stores.

5. DISCUSSIONS

The first stage begins by forming a k-1 itemset candidate from 181 item sales data with a total of 65 transactions. The author provides a minimum support value limitation of 0.3 or 30% and a minimum confidence of 0.65 or 65%. The support value for the formation of k-1 itemset is calculated using the following formula (1).

Table 2
Candidate K-1 Item set

No	Item set	Qty	Calculation	Support %	Description
1	Cooking Spices	8	$(8/65)*100\%$	12,3	Fails
2	Biscuits	49	$(49/65)*100\%$	75,4	Meets
3	Wafer	44	$(44/65)*100\%$	67,7	Meets
4	Cereal	11	$(11/65)*100\%$	16,9	Fails
5	Deterjen	4	$(4/65)*100\%$	6,2	Fails
6	Personal Care	22	$(22/65)*100\%$	33,8	Meets
7	Drinks	20	$(20/65)*100\%$	30,8	Meets
8	Frozen Food	11	$(11/65)*100\%$	16,9	Fails
9	Dessert	12	$(12/65)*100\%$	18,5	Fails

* Corresponding author



From the process of the first stage of forming the K-1 itemset above, 4 itemsets are obtained that meet the minimum support value of more than 30%, namely biscuits, wafers, personal care, and drinks. Then the formation of a combination of 2 items with a minimum support of 30% can be solved by the formula (2).

After that, at the 2nd iteration, the calculation results for each combination of k-2 item sets can be seen in the following table:

Table 3
Candidate K-2 Item set

No	Item set	Qty	Calculation	Support %	Description
1	Biscuits, Wafer	29	(29/65)*100%	44,6	Meets
2	Biscuits, Personal Care	20	(20/65)*100%	30,8	Meets
3	Biscuits, Minuman	10	(10/65)*100%	15,4	Fails
5	Wafer, Personal Care	6	(6/65)*100%	9,2	Fails
6	Wafer, Drinks	13	(13/65)*100%	20,0	Fails
7	Personal Care, Drinks	4	(4/65)*100%	6,2	Fails

Here are the item sets that meet the minimum support:

Table 4
Entire Item set Iteration Results

Item set	Qty	Support %
Biscuits, Wafer	29	44,6
Biscuits, Personal Care	20	30,8

After all item sets are formed, the next step is to separate the antecedent from the consequent to determine all possible associations. Then further in determining the association rules, the confidence value of the combination will be calculated, provided that the minimum confidence value is 65% using the following formula (3).

Table 5
Confidence Calculation

Antecedent	Consequent	Calculation	Confidence %	Description
Biscuits	Wafer	(29/49)*100	59,2	Fails
Wafer	Biscuits	(29/44)*100	65,9	Meets
Biscuits	Personal Care	(20/49)*100	40,8	Fails
Personal Care	Biscuits	(20/22)*100	90,9	Meets

for item sets that are worth less than 65% will be eliminated, with these results an expected association rule can be formed. Here is the item set that has been formed with a minimum confidence of 65%.

* Corresponding author



Table 6
Item set meets minimum confidence

Antecedent	Consequent	Calculation	Confidence %	Description
Wafer	Biscuits	$(29/44)*100$	65,9	Meets
Personal Care	Biscuits	$(20/22)*100$	90,9	Meets

After the item set is determined, the last step is to validate the accuracy of the rules obtained by calculating the lift ratio. The lift ratio calculation starts with finding the benchmark confidence value. Here is the formula for calculating benchmark confidence (5):

Table 7
Benchmark Confidence Calculation

Antecedent	Consequent	Calculation	Benchmark Confidence %
Wafer	Biscuits	$(49/65)*100\%$	75,4
Personal Care	Biscuits	$(49/65)*100\%$	75,4

After getting the benchmark confidence value, then validate the rules obtained by calculating the lift ratio. Here is the formula for calculating the lift ratio :

Table 8
Lift Ratio Calculation

Antecedent	Consequent	Confidence %	Benchmark Confidence %	Lift Ratio
Wafer	Biscuits	65,9	75,4	0,874
Personal Care	Biscuits	90,9	75,4	1,206

Based on the analysis conducted, it was found that when "Wafer" is the antecedent, "Biscuit" as the consequent results in a lift ratio of 0.874. This lift ratio is less than 1, which indicates the relationship between "Wafer" and "Biscuit" is invalid. In addition, when "Personal Care" is the antecedent, "Biscuits" as the consequent results in a lift ratio of 1.206. This lift ratio is also more than 1, indicating the relationship between "Personal Care" and "Biscuits" is valid.

6. CONCLUSION

Based on the results of the application of association rules using the apriori algorithm on Intan Store sales transaction data using a minimum support of 30% and a minimum confidence of 65%, consumer purchasing patterns can be seen from the combination of sales of products purchased, the association pattern formed is if consumers buy Personal Care then consumers will also buy Biscuits with a support of 30.8% and a confidence of 90.9% with a validation lift ratio of 1.206.

7. REFERENCES

- A Fadila Shely, S Setiawansyah, & D Dedi. (2021). ANALISIS DATA PENJUALAN HANDPHONE DAN ELEKTRONIK MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS : CV REY GASENDRA). *Jurnal Telematics and Information Technology*, 2(1), 1–6.
- Agustiani, N., Suhendro, D., Saputra, W., & Tunas Bangsa Pematangsiantar, S. (2020). Penerapan Data Mining Metode Apriori Dalam Implementasi Penjualan Di Alfamart. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, 300–304.
- Annisa Ilham Fathimah, Sudin Saepudin, & Edwinanto. (2022). PENERAPAN DATA MINING DENGAN METODE APRIORI PADA PENJUALAN SEMBAKO (STUDI KASUS: GROSIR SEMBAKO LINA). *Jurnal Rekayasa*

* Corresponding author



- Teknologi Nusa Putra*, 8(2), 123–131. <https://doi.org/10.52005/rekayasa.v8i2.271>
- Faqih, A. S., & Wahyudi, A. D. (2022). Rancang Bangun Sistem Informasi Penjualan Berbasis Web (Studi Kasus : Matchmaker). *Jurnal Teknologi Dan Sistem Informasi*, 3(2), 1–8. <http://jim.teknokrat.ac.id/index.php/JTSI>
- H Arva Abhyoso, A Rachmadita, & D Irfan. (2020). Implementasi Metode Algoritma Apriori Untuk Menemukan Association Rules Data Obat Di Rumah Sakit Port Medical Center Jakarta. *E-Proceeding of Engineering*, 7(2), 7096.
- Khanza, M., Toyib, R. T., & Onsardi, O. (2021). Implmentasi Algoritma Apriori Untuk Meningkatkan Penjualan Handphone di Toko Mardha Cell. *JSAI (Journal Scientific and Applied Informatics)*, 4(2), 221–235. <https://doi.org/10.36085/jsai.v4i2.1674>
- Kristania, Y. M., & Listanto, S. (2022). IMPLEMENTASI DATA MINING TERHADAP DATA PENJUALAN DENGAN ALGORITMA APRIORI PADA PT. DUTA KENCANA SWAGUNA. *Jurnal Teknoinfo*, 16(2), 364. <https://doi.org/10.33365/jti.v16i2.1973>
- Lienata, B., Fenriana, I., Andre, A., & Safitri, R. D. (2021). Penerapan Data Mining Pada Penjualan Pakaian Brand Expand Dengan Algoritma Apriori Menggunakan Metode Association Rules PT. Vidiaelok Lestari Garmino. *Algor*, 3(1), 83–95. <https://doi.org/10.31253/algor.v3i1.647>
- M Nanda Nurisya, K Nur Isnaeni, W Neta Tri, T Latifah Adi, & S Pungkas. (2022). Penerapan Algoritma Apriori Pada Transaksi Penjualan Untuk Rekomendasi Menu Makanan Dan Minuman . *JURNAL NASIONAL TEKNOLOGI DAN SISTEM INFORMASI*, 08(01), 009–016.
- Prasetyo, A., Sastra, R., & Musyaffa, N. (2020). IMPLEMENTASI DATA MINING UNTUK ANALISIS DATA PENJUALAN DENGAN MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS DAPOERIN’S). *Jurnal Khatulistiwa Informatika*, 8(2). <https://doi.org/10.31294/jki.v8i2.8994>
- Riyanto, A., & Susanti, M. (2020). Analisa Data Mining Terhadap Penjualan Food Dengan Metode Apriori Pada Kopsyahira. *Bianglala Informatika*, 8(1), 44–48. <https://doi.org/10.31294/bi.v8i1.8148>
- Saefudin, & F Donny. (2020). PENERAPAN DATA MINING REKOMENDASI BUKU MENGGUNAKAN ALGORITMA APRIORI. *Jurnal Sistem Informasi*, 7(1), 50–56.
- Sahara, W., Saragih, S. D., & Windarto, A. P. (2022). Teknik Asosiasi Datamining Dalam Menentukan Pola Penjualan dengan Metode Apriori. *TIN: Terapan Informatika Nusantara*, 2(12), 684–689. <https://doi.org/10.47065/tin.v2i12.1577>
- Syabania, R., & Rosmawarni, N. (2021). PERANCANGAN APLIKASI CUSTOMER RELATIONSHIP MANAGEMENT (CRM) PADA PENJUALAN BARANG PRE-ORDER BERBASIS WEBSITE. In *Jurnal Rekayasa Informatika* (Vol. 10, Issue 1).
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>
- Triayudi, A. (2022). Penerapan Algoritma Apriori Data Mining Untuk Menentukan Penyusunan Layout Barang Pada Toko Ritel. *Building of Informatics, Technology and Science (BITS)*, 4(2). <https://doi.org/10.47065/bits.v4i2.2303>
- Triayudi, A., & Iskandar, A. (2022). Penerapan Data Mining Dalam Penentuan Prioritas Pemesanan Produk Berdasarkan dengan Data Penjualan Barang Menggunakan Algoritma Apriori. *Journal of Computer System and Informatics (JoSYC)*, 4(1), 25–30. <https://doi.org/10.47065/josyc.v4i1.2523>
- Triyuliani, N., Syahril, M., & Syahputra, Y. H. (2020). Analisa Association Rules Untuk Pola Penjualan Cake Pada Ratu Bakery Menggunakan Algoritma Apriori. *Jurnal CyberTech*, 3(9), 1433–1444. <https://ojs.trigunadharma.ac.id/>

* Corresponding author

