The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 1, February 2025
https://doi.org/10.33022/ijcs.v14i1.4652

# Implementation of Data Mining to Analyze Consumer Purchasing Patterns Using the Apriori Algorithm

**Arief Sulistyo Wibowo[1], Dwi Sukma Donoriyanto[2]**
ariefsulistyowibowo20@gmail.com[1], dwisukama.ti@upnjatim.ac.id[2]
[1,2]Department of Industrial Engineering, Pembangunan Nasional Veteran Jawa Timur University

**Article Information**

Received : 16 Jan 2025
Revised : 6 Feb 2025
Accepted : 28 Feb 2025

**Keywords**

Apriori, Association Rules, Chemical Product, Data Mining

**Abstract**

CV XYZ is a retail store located on Jl. Tidar, Surabaya, East Java, specializing in the trade of industrial and household chemical products. To remain competitive amid rapid technological advancements and increasing competition in the area, CV XYZ has adopted online sales through the Shopee platform. This study aims to analyze consumer purchase patterns using the Apriori algorithm based on sales transaction data to support effective and efficient marketing strategies. Based on the analysis, five purchasing patterns were identified: (1) if Silicon Oil is purchased, Counterdust will also be purchased with a support value of 0.391 and a confidence value of 0.818; (2) if Cocamidopropyl Betain is purchased, Counterdust will also be purchased with a support value of 0.391 and a confidence value of 0.857; (3) if Car Shampoo is purchased, Counterdust will also be purchased with a support value of 0.359 and a confidence value of 0.868; (4) if Talc is purchased, Counterdust will also be purchased with a support value of 0.348 and a confidence value of 0.842; and (5) if Linear Alkylbenzene Sulfonate is purchased, Counterdust will also be purchased with a support value of 0.359 and a confidence value of 0.892. These findings indicate that data mining techniques using the Apriori algorithm provide an effective approach for identifying consumer purchasing patterns of chemical products sold online through the Shopee platform. This insight can help businesses optimize their marketing strategies and decision-making processes.

Attribution-ShareAlike 4.0 International License

## A. Introduction

Competition in the retail business today compels companies to continuously develop various strategies and innovations to enhance revenue performance, as reflected in product sales [1]. Achieving success in this business requires a deep understanding of consumer behavior and their shopping preferences. Comprehensive knowledge of consumer needs, purchasing timing, and frequently co-purchased products is essential for designing effective marketing strategies and efficient inventory management [2]. By understanding consumer shopping patterns, retail businesses can formulate more effective and targeted marketing strategies.

One method to identify consumer purchasing patterns is through the use of sales transaction data. Sales transaction data can be reanalyzed to uncover consumer purchasing patterns [3]. Essentially, sales transaction data is an asset that can be leveraged to optimize more targeted business strategies. Transaction data from each consumer purchase can be transformed into a dataset, which, when processed effectively, can provide valuable information to support strategic business decision-making [4].

CV XYZ is a retail store located in Surabaya, East Java. The store operates in the retail trade industry, specializing in industrial and household chemical products. With technological advancements and increasing competition in Jl. Tidar, known as an area with numerous chemical product stores, CV XYZ has also adapted by conducting online sales through the e-commerce platform Shopee. Sales transaction data from CV XYZ on Shopee has been well-recorded by the platform's system. However, the transaction data, comprising 2,251 records from July, August, and September, is currently utilized only as an archive and has not been analyzed further. This data could be processed to generate valuable insights, particularly in understanding consumer purchasing patterns.

To utilize this data, an analytical method capable of uncovering hidden patterns within the data is required. One technique that can be applied is data mining. Data mining is a series of processes used to explore databases in order to discover new knowledge that cannot be identified manually [5]. It is a process of uncovering meaningful relationships, patterns, and trends by examining large datasets stored in databases using pattern recognition techniques [6]. Data mining can also be employed to identify purchasing patterns from customers when buying combinations of items using specific methods [7]. In this research, the data mining technique applied is the association rule technique.

Association analysis or association rules is a technique used to identify associative relationships between items and to discover items that frequently appear together at the same time [8]. The Association Rules method has two objectives: to determine which products are typically sold together and to identify the rules underlying these associations [9] One approach to establishing association rules in data mining is by using the Apriori algorithm. This algorithm aims to uncover interesting relationships or associations between items in large datasets, particularly within transactional databases. The Apriori algorithm is highly efficient and effective in identifying itemsets that frequently appear together in transactions [10].

Several previous studies have applied the Apriori algorithm in purchasing pattern analysis across various sectors. For instance, a study by Santoso titled "Application of the Association Rule Method Using the Apriori Algorithm to Identify Sales Patterns: A Case Study at Indomaret Tanjung Anom" demonstrated that the Apriori algorithm can be effectively applied to identify purchasing transaction patterns by combining products into sales packages or bundling strategies to enhance product sales [11]. Meanwhile, a study by Purwati titled "Comparison of the Apriori and FP-Growth Data Mining Methods for Identifying Sales Patterns" focused on perfume products in e-commerce, with the analysis results indicating that association rules generated using the Apriori algorithm were highly suitable for application [12].

However, research related to industrial and household chemical products remains limited, making this study a unique contribution in this context. This research contributes to the development of a data mining approach using the Apriori algorithm to analyze purchasing patterns in the context of e-commerce for chemical products. Unlike previous studies that focused on general retail or fast-moving consumer goods, this study emphasizes the effective application of the Apriori algorithm to industrial and household chemical products, which exhibit less frequent purchasing patterns.

Therefore, this research aims to analyze consumer purchasing patterns based on association rules using sales transaction data. The insights gained from analyzing consumer purchasing patterns are expected to assist the company in making informed decisions to enhance effective and efficient marketing strategies.

## B. Research Method

This research was conducted in September 2024 until the data was met. The stages of this research are:

### 1) Identification and Operational Definition of Variables

This research variable uses a Dependent variable with Independent variable.

a. Dependent Variable

The dependent variable is the variable that is influenced or affected by the independent variable. In this study, the dependent variable is the consumer purchasing patterns. These purchasing patterns are represented by the association rules generated through the Apriori algorithm.

b. Independent Variable

The independent variable is the variable that influences the dependent variable. In this study, the independent variable consists of the products sold in online transactions on the Shopee e-commerce platform.

### 2) Problem Solving Steps

The problem solving stages of this research are as follows:

a. Field Study

The field study is an observational phase conducted directly in the field to obtain detailed insights into the issues being investigated based on existing conditions.

b. Literature Review
The literature review is a phase aimed at fulfilling theoretical needs related to the researched issues, with the goal of identifying a framework applicable to solving the problem. The literature utilized can come from books, research journals, and previous studies.
c. Problem Formulation
This phase involves developing the background by identifying the problems to be addressed, specifically how to implement data mining to analyze consumer purchasing patterns using the Apriori algorithm.
d. Research Objectives
This phase involves determining the objectives to be achieved in resolving the issues. The objective of this study is to analyze consumer purchasing patterns based on sales transaction data using the Apriori algorithm.
e. Variable Identification
This phase identifies the variables involved in the study, including dependent and independent variables.
f. Data Collection
Data collection is the phase of gathering supporting information to address the research problems. The data collection technique used is secondary data collection in the form of archived online sales transaction data from Shopee during July, August, and September, sourced from the official account of CV XYZ.
g. Pre-Processing
This critical phase ensures the data quality aligns with the requirements of the Apriori algorithm. Pre-processing steps include Data Selection, Data Cleaning, and Data Transformation.
h. Frequent Itemset Formation
This phase involves analyzing high-frequency patterns, identifying item combinations that meet the minimum support value requirements within the dataset.
i. Association Rule Formation
After identifying all high-frequency patterns, associative rules are derived that meet the minimum confidence value requirements.
j. Results and Discussion
The results and discussion section forms the core of a scientific report. This phase analyzes and discusses the results and insights gained from the association rules formed.
k. Conclusion and Recommendations
The conclusion and recommendations provide a brief, clear, and systematic summary of the overall analysis results, along with suggestions or opinions from the researcher regarding problem-solving based on the intended objectives.

**3) Data Collection Method**

Data collection is a phase aimed at gathering supporting information for the research, which will be utilized to address the research problems. The data

collection technique employed is secondary data collection, consisting of online sales transaction data from the official Shopee account of CV XYZ.

.

## C. Result and Discussion

### 1) Data Collection

The dataset used in this study comprises sales transaction data from the official Shopee account of CV XYZ, with a total of 2,251 transaction records collected during the period of July, August, and September 2024 with a total of 105 different products sold.. This dataset is considered sufficient and representative due to its adequate transaction volume and product diversity, both of which are essential for identifying meaningful purchasing patterns. The large number of transactions recorded over a three-month period reflects purchasing behavior across different time frames.

Previous studies on consumer purchasing patterns using the Apriori algorithm have shown that datasets of similar or even smaller sizes can yield valuable association rules. For instance, A. R. Efrat et al. analyzed the implementation of data mining with the Apriori algorithm to determine consumer purchasing patterns in a minimarket with a smaller data volume and fewer product variations sold, successfully identifying significant trends in consumer behavior [13].

Furthermore, this dataset includes 105 unique products, ensuring a wide range of purchase combinations. Greater product diversity increases the likelihood of discovering meaningful association rules, as it allows for the identification of frequently co-purchased items across various purchasing contexts.

**Table 1.** The transaction data from July, August, and September.

| No | Order Number | Order Status | Cancellation Reason | Cancellation Status | ... | Province | Order Completion Time |
|---|---|---|---|---|---|---|---|
| 1 | 240701EVCPCAY0 | Completed | | | ... | YOGYAKARTA | 2024-07-03 09:51 |
| 2 | 240701EW1TK5GT | Completed | | | ... | EAST JAVA | 2024-07-03 07:31 |
| 3 | 240701EY866J52 | Completed | | | ... | EAST JAVA | 2024-07-04 02:36 |
| 4 | 240701EY866J52 | Completed | | | ... | EAST JAVA | 2024-07-04 02:36 |
| 5 | 240701F364DAWH | Completed | | | ... | EAST JAVA | 2024-07-04 05:05 |
| 6 | 240701FEAT0RAQ | Completed | | | ... | EAST JAVA | 2024-07-07 06:41 |
| 7 | 240701FK8D0DJ0 | Completed | | | ... | EAST JAVA | 2024-07-05 04:20 |
| 8 | 240701FK8D0DJ0 | Completed | | | ... | EAST JAVA | 2024-07-05 04:20 |
| 9 | 240701FK8D0DJ0 | Completed | | | ... | EAST JAVA | 2024-07-05 04:20 |
| 10 | 240701FK8D0DJ0 | Completed | | | ... | EAST JAVA | 2024-07-05 04:20 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 2249 | 240930BNJFTYEU | Completed | | | | EAST JAVA | 2024-10-02 17:08 |

| No | Order Number | Order Status | Cancellation Reason | Cancellation Status | ... | Province | Order Completion Time |
|---|---|---|---|---|---|---|---|
| 2250 | 240930BWKQ08PH | In Transit | | | ... | EAST JAVA | |

Source : Company Data (2024)

The sales transaction data consists of 49 columns, including "Order Number," "Order Status," "Cancellation Reason," "Cancellation Status," "Tracking Number," "Shipping Option," "Delivery to Counter/Pick-up," "Order Must Be Shipped By," "Shipping Time Set," "Order Creation Time," "Payment Time," "Payment Method," "Parent SKU," "Product Name," "SKU Reference Number," "Variation Name," "Original Price," "Price After Discount," "Quantity," "Returned Quantity," "Total Product Price," "Total Discount," "Seller Discount," "Shopee Discount," "Product Weight," "Quantity Ordered," "Total Weight," "Seller-Covered Voucher," "Coin Cashback," "Shopee-Covered Voucher," "Discount Bundle," "Discount Bundle (Shopee Discount)," "Discount Bundle (Seller Discount)," "Shopee Coin Deduction," "Credit Card Discount," "Shipping Cost Paid by Buyer," "Estimated Shipping Cost Deduction," "Return Shipping Cost," "Total Payment," "Estimated Shipping Cost," "Buyer Notes," "Notes," "Username (Buyer)," "Recipient Name," "Phone Number," "Shipping Address," "City/Regency," "Province," and "Order Completion Time,".

The sales transaction data will proceed to the preprocessing phase, which will be performed using the Anaconda software with a Jupyter Lab environment and Python language. In the preprocessing process, steps such as data selection, data cleaning, and data transformation will be carried out. Subsequently, within the sales transaction data spanning three months, the sold products will undergo name adjustments to facilitate the analysis process.

.

**2) Data Processing**

a. Data Selection

In this phase, data selection is performed, where from the dataset consisting of 49 columns, only the relevant columns required for the subsequent processes are chosen. The selected columns in this process are "Order Status" with the value "Completed," "Payment Time," and "Product Name."

b. Data Cleaning

In the data cleaning phase, the data is cleaned by removing duplicates. Since the "Product Name" column can be represented by the product value for each payment time, duplicate entries are eliminated. Next, row aggregation is performed based on the "Payment Time" column. This step is necessary because chemical products are typically purchased individually in each transaction, requiring aggregation by payment time to examine item combinations within a single day over the 3-month period

c. Data Transformation

Following the data cleaning, the data undergoes transformation, converting it into a tabular format with binary values (True or False). A value of 1 represents "True" or "purchased," while a value of 0 indicates "False" or "not purchased." Data transformation is essential as the Apriori algorithm, with the help of software, requires the data to be in this format.

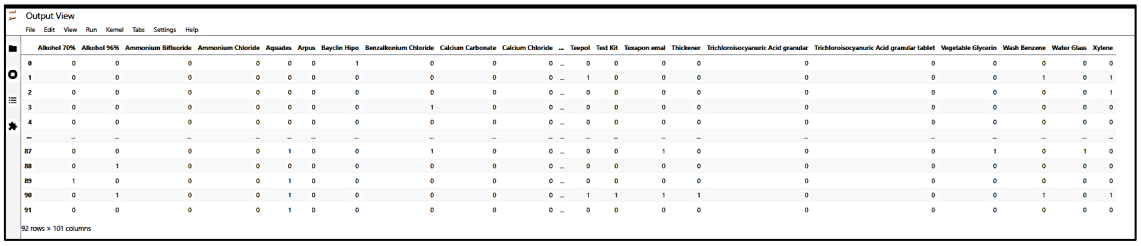


**Figure 1.** Data Transformation

Upon the completion of the transformation process, the data preprocessing phase is considered finished, and the data is now ready to proceed to the next stage, which involves the implementation of the Apriori algorithm.

d. Formation of Frequent Itemsets

After data transformation, the next step in the Apriori algorithm is the formation of frequent itemsets. This phase involves identifying combinations of items that frequently occur together, and it is carried out with the assistance of the Orange Data Mining software. During this process, the parameters of support and confidence are defined. The significance of an association rule can be measured using two parameters: support (support value), which indicates the percentage of the combination of these items in the database, and confidence (certainty value), which reflects the strength of the relationship between the items in the association rule [14].

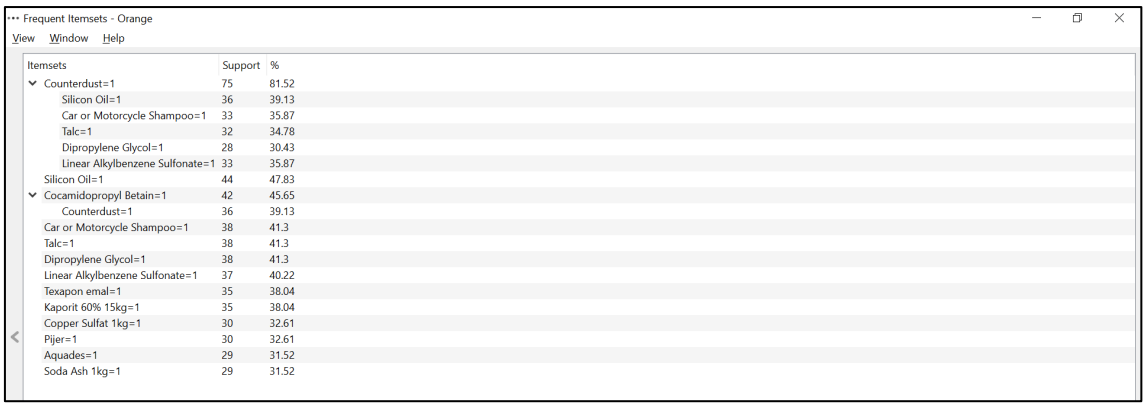


**Figure 3.** Formation of Frequent Itemsets

Based on the results of the frequent itemset analysis with a minimum support threshold of 30%. Support is defined as the probability of consumers purchasing multiple products together across all transactions. The support of an association rule represents the percentage of occurrences of a specific combination of items in the database. For instance, when considering items A and B, support refers to the proportion of transactions in the database that contain both items [15]. In this analysis, a total of 13 products were identified with a 1-itemset, while 6 products were found with a 2-itemset.

**Table 2.** Formation of 1 Itemsets

| No | Product | Number of Transactions | Support (%) |
|---|---|---|---|
| 1 | Counterdust | 75 | 81.52 |
| 2 | Silicon Oil | 44 | 47.83 |
| 3 | Cocamidopropyl Betain | 42 | 45.65 |
| 4 | Car or Motorcycle Shampoo | 38 | 41.3 |
| 5 | Talc | 38 | 41.3 |
| 6 | Dipropylene Glycol | 38 | 41.3 |
| 7 | Linear Alkylbenzene Sulfonate | 37 | 40.22 |
| 8 | Texapon emal | 35 | 38.04 |
| 9 | Kaporit 60% 15kg | 35 | 38.04 |
| 10 | Copper Sulfat 1kg | 30 | 32.61 |
| 11 | Pijer | 30 | 32.61 |
| 12 | Aquades | 29 | 31.52 |
| 13 | Soda Ash 1kg | 29 | 31.52 |

**Table 3.** Formation of 2 Itemsets

| No | Product | Number of Transactions | Support (%) |
|---|---|---|---|
| 1 | Counterdust - Silicon Oil | 36 | 39.13 |
| 2 | Counterdust - Car or Motorcycle Shampoo | 33 | 35.87 |
| 3 | Counterdust - Talc | 32 | 34.78 |
| 4 | Counterdust - Dipropylene Glycol | 28 | 30.43 |
| 5 | Counterdust - Linear Alkylbenzene Sulfonate | 33 | 35.87 |
| 6 | Cocamidopropyl Betain - Counterdust | 36 | 39.13 |

e. Formation of Association Rule

Subsequently, after the formation of frequent itemsets, the next step is the generation of association rules based on the frequent itemsets

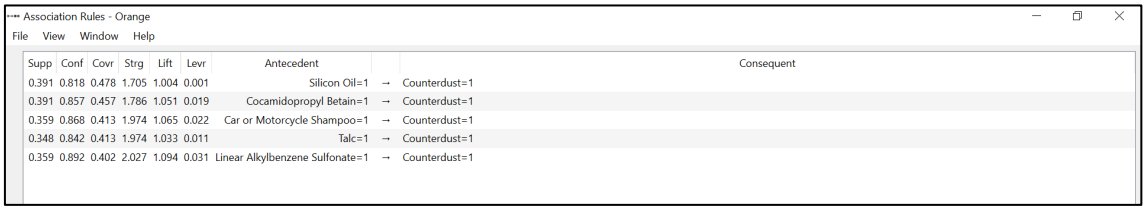


**Figure 3.** Formation of Association Rules

Based on the results of the analysis, association rules were generated with a minimum support threshold of 30% and a minimum confidence threshold of 80%. The determination of confidence was carried out to find the certainty of the truth of itemset purchases that occurred in the transaction dataset. Confidence is calculated by dividing the total itemsets purchased together by the total number of transactions that contain the antecedent items in the dataset [16]. The analysis resulted in a total of 5 association rules.

**Table 4.** Final of Association Rule

| No | Product | Support | Confidence |
|---|---|---|---|
| 1 | Silicon Oil - Counterdust | 0.391 | 0.818 |
| 2 | Cocamidopropyl Betain - Counterdust | 0.391 | 0.857 |
| 3 | Car or Motorcycle Shampoo - Counterdust | 0.359 | 0.868 |
| 4 | Talc - Counterdust | 0.348 | 0.842 |
| 5 | Linear Alkylbenzene Sulfonate - Counterdust | 0.359 | 0.892 |

Based on the table above, a total of 5 purchasing patterns were identified. These include: if the product Silicon Oil is purchased, Counterdust will also be purchased with a support value of 0.391 and a confidence value of 0.818; if the product Cocamidopropyl Betain is purchased, Counterdust will also be purchased with a support value of 0.391 and a confidence value of 0.857; if the product Car Shampoo is purchased, Counterdust will also be purchased with a support value of 0.359 and a confidence value of 0.868; if the product Talc is purchased, Counterdust will also be purchased with a support value of 0.348 and a confidence value of 0.842; and if the product Linear Alkylbenzene Sulfonate is purchased, Counterdust will also be purchased with a support value of 0.359 and a confidence value of 0.892. Therefore, it can be concluded that the data mining technique using the Apriori algorithm can be an effective solution for determining consumer purchasing patterns for chemical products sold online through the Shopee platform.

## D. Conclusion

Based on the analysis and processing of data from online sales transactions through the official Shopee platform account of CV XYZ, utilizing data mining techniques with the Apriori algorithm, a minimum support of 30%, and a minimum confidence of 80%, conducted with the assistance of Orange Data Mining software, the following conclusions were drawn: Five consumer purchasing patterns were identified. These patterns include: if a consumer purchases Silicon Oil, they are likely to also purchase Counterdust, with a support value of 0.391 and a confidence value of 0.818; if the product Cocamidopropyl Betain is purchased, Counterdust will also be purchased with a support value of 0.391 and a confidence value of 0.857; if the product Car Shampoo is purchased, Counterdust will also be purchased with a support value of 0.359 and a confidence value of 0.868; if the product Talc is purchased, Counterdust will also be purchased with a support value of 0.348 and a confidence value of 0.842; and if Linear Alkylbenzene Sulfonate is purchased, Counterdust will also be purchased with a support value of 0.359 and a confidence value of 0.892.

This indicates that customers tend to purchase these products together, which can serve as the basis for bundling strategies or discount promotions. Additionally, the high confidence levels in the association rules suggest consistent purchasing patterns, which can be utilized in product recommendation systems for new customers. The findings of this study provide insights into consumer purchasing patterns and their impact on business strategies, particularly in inventory management and product promotions. By implementing the Apriori algorithm, companies can identify frequently purchased product combinations and apply data-driven marketing strategies.

## E. References


[1] F. Achmad, O. Nurdiawan, and Y. Arie Wijaya, "Analysis of Consumer Purchase Transaction Patterns in a Health Retail Store Using the FP-Growth Algorithm," JATI (J. of Tech. Inf. Students), vol. 7, no. 1, pp. 168–175, 2023, doi: 10.36040/jati.v7i1.6210.

[2] G. Soepriyono and A. Triayudi, "Implementation of Data Mining with the Apriori Algorithm in Determining Laptop Accessories Purchase Patterns," J. of Media Inform. Budidarma, vol. 7, no. 4, p. 2087, 2023, doi: 10.30865/mib.v7i4.6555.

[3] E. Munanda and S. Monalisa, "Application of the FP-Growth Algorithm in Sales Transaction Data for Layout Determination," J. of Scientific Engineering and Management Systems, vol. 7, no. 2, pp. 173–184, 2021, [Online]. Available: http://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/13253.

[4] V. C. Nisa and F. N. Khasanah, "Apriori Algorithm in Identifying Consumer Purchase Patterns for Beverage Products," INFORMAL Informatics J., vol. 8, no. 2, p. 156, 2023, doi: 10.19184/isj.v8i2.41514.

[5] G. W. Aji and P. A. R. Devi, "Data Mining Implementation for Product Transaction Patterns Using the Apriori Method," Sinkron, vol. 8, no. 1, pp. 421–432, 2023, doi: 10.33395/sinkron.v8i1.12071.

[6] A. N. Rahmi and Y. A. Mikola, "Implementation of the Apriori Algorithm to Determine Purchase Patterns in Customers (Case Study: Bakoel Sembako Store)," Inf. Systems J., vol. 4, no. 1, pp. 14–19, 2021, [Online]. Available: https://jurnal.amikom.ac.id/index.php/infos/article/view/561.

[7] A. M. Ugrasena and A. Z. Falani, "Utilization of the FP-Growth Algorithm to Determine Sales Strategies at Teras Garden Coffee Shop," Syst. Inf. Syst. Informatics J., vol. 6, no. 2, pp. 33–37, 2021, doi: 10.29080/systemic.v6i2.977.

[8] A. Novianti and E. Elisa, "Determination of Association Rules in Purchase Patterns at Minimarkets Using the Apriori Algorithm," Build. Informatics, Technol. Sci., vol. 2, no. 1, pp. 64–70, 2020, [Online]. Available: http://ejurnal.seminar-id.com/index.php/bits/article/view/300.

[9] N. L. W. S. R. Ginantra et al., Data Mining and Algorithm Application, 2021.

[10] E. Omol, D. Onyango, L. Mburu, and P. Abuonji, "Apriori Algorithm and Market Basket Analysis to Uncover Consumer Buying Patterns: Case of a Kenyan Supermarket," vol. 5, no. 2, pp. 51–63, 2024.

[11] M. H. Santoso, "Application of the Association Rule Method Using the Apriori Algorithm to Find Sales Patterns: Case Study of Indomaret Tanjung Anom," Brill. Res. Artif. Intell., vol. 1, no. 2, pp. 54–66, 2021, doi: 10.47709/brilliance.v1i2.1228.

[12] N. Purwati, Y. Pedliyansah, H. Kurniawan, S. Karnila, and R. Herwanto, "Comparison of the Apriori and FP-Growth Data Mining Methods to Determine Sales Patterns," J. Inform. J. of IT Development, vol. 8, no. 2, pp. 155–161, 2023, doi: 10.30591/jpit.v8i2.4876.

[13] A. R. Efrat, R. Gernowo, and Farikhin, "Consumer Purchase Patterns Based on Market Basket Analysis Using the Apriori Algorithm," J. Phys. Conf. Ser., vol. 1524, no. 1, 2020, doi: 10.1088/1742-6596/1524/1/012109.

[14] R. Alfianzah, R. I. Handayani, and M. Murniyati, "Implementation of the Apriori

Algorithm Data Mining to Increase Sales," SinkrOn, vol. 5, no. 1, p. 17, 2020, doi: 10.33395/sinkron.v5i1.10587.

[15] M. Sholihul, I. Gede, P. Wirarama, W. Wirawan, and F. Bimantoro, "Implementation of the Apriori Algorithm for Sales Transaction Analysis in Pharmacy (Case Study: Gilda Farma Pharmacy 2)," J. of Info. Tech., Comput., and Apps., vol. 4, no. 1, pp. 63–74, 2022, [Online]. Available: http://jtika.if.unram.ac.id/index.php/JTIKA/.

[16] R. Afandi, R. Martiansyah, L. M. Sari, S. F. Intan, and J. Pulungan, "Application of the Apriori Algorithm to Determine the Purchase Transaction Pattern for Drinks and Food at Coffee Gubuk," vol. 1, no. 2, pp. 111–118, 2022.