**NEWTON: Networking and Information Technology**
Vol. 4 No. 2 October 2024, Page. 105 - 114
E-ISSN: 2797-0728

# Implementation of Apriori and Fp-Growth Algorithms In Forming Association Patterns Based On Unwaha Cooperative Sales Transactions

**Achmad Agus Athok Miftachuddin[1], Dhita Cahyaningtyas[2]**
[1]Departemen of Information Systems, KH. A. Wahab Hasbullah
[2]Departemen of Information Systems, KH. A. Wahab Hasbullah

*Email: agusathok@unwaha.ac.id, cahyaningtyasdhita@gmail.com

## ABSTRACT
*This study implements the Apriori and FP-Growth algorithms to identify association rules in sales transaction data from the UNWAHA Multi-Purpose Cooperative. Both algorithms successfully discovered product relationships, with similarities in rules for items like MAKARONI ASEP, KRUPUK PAK JONO, and KRUPUK 500. The FP-Growth algorithm, implemented using RapidMiner, outperformed Apriori in processing speed by 11 seconds and demonstrated higher accuracy in rule generation. Optimal results were achieved with minimum support and confidence values of 0.3 and 0.9 for Apriori (generating 5 rules), and 0.52 and 0.9 for FP-Growth (generating 6 rules). These settings balanced between generating too many rules, which could complicate interpretation, and too few, which might miss important patterns. Based on the analysis, strategic recommendations for the cooperative include implementing product bundling and discounts for frequently co-purchased items nearing expiration, optimizing product placement by grouping commonly associated items (e.g., MAKARONI ASEP, KRUPUK PAK JONO, KRUPUK 500, SOSIS SO NICE, and YUPI ALL VARIAN 5G) in easily accessible locations, and increasing stock for high-demand products like LE MINERALE. This research demonstrates the practical application of association rule mining in retail, offering data-driven insights to enhance sales strategies and inventory management for the UNWAHA Cooperative.*

***Keywords**: Apriori; FP-Growth; Association Rules; RapidMiner.*

## INTRODUCTION
The rapid development of information technology has opened up opportunities for business entities to optimize their transaction data more effectively. One way to gain valuable insights from transaction data is through association pattern analysis. In this context, cooperatives as member-based economic institutions can also benefit from applying data mining techniques to analyze their members' purchasing patterns. This research focuses on implementing two popular algorithms in forming association patterns, namely Apriori and FP-Growth, on the sales transaction data of UNWAHA Cooperative. These two algorithms were chosen for their ability to identify relationships between items in purchase transactions, which can provide valuable information for cooperative management in strategic decision-making.

Apriori and FP-Growth have different approaches to analyzing transaction data. Apriori uses the generate-and-test method to find frequent itemsets, while FP-Growth uses a tree data structure to compress transaction information. By comparing the results of these two algorithms, this study aims to provide a more comprehensive understanding of association patterns in UNWAHA Cooperative's sales transactions. The results of this research are expected to assist cooperative management in optimizing marketing strategies, inventory management, and improving services to members. Additionally, this study contributes to the literature on the application of data mining techniques in the context of cooperatives in Indonesia.

# METHOD

## Cooperative

A cooperative is a people's economic organization with a social character, consisting of individuals or cooperative legal entities that form an economic structure as a joint effort based on the principle of kinship (Widayu et al., 2017).

## Data Mining

The public definition of data mining is a method of searching for hidden knowledge patterns that were previously unknown from a very large set of data in databases, data warehouses, or other storage media. Data mining is used to extract added value in the form of information that is not known manually from a database. Information is obtained by extracting and recognizing important or interesting patterns from data contained in the database (Harahap & Sulindawaty, 2020).

## Association Rules

In the field of data mining science, there is a method called association rule. This method is often also called market basket analysis. Association rule mining is a procedure for finding relationships between items in a specified data set. Association rules include two stages (Rerung, 2018):

a. Analysis of high-frequency patterns
   Support value is the percentage of records that contain the combination.

$$Support\ (A) = \frac{\sum Transactions\ Containing\ (A) \times 100}{Total\ Transactions\ (A)}$$

b. Formation of Association Rules
   The accuracy of an association rule is often called confidence. Confidence is the strength of the relationship between items in an associative rule.

$$Confidence\ (A => B) = \frac{Support\ (A\ and\ B) \times 100}{Support\ (A)}$$

## Apriori Algorithm

The Apriori algorithm is a type of association rule in data mining that explains the association of several attributes often called affinity analysis or market basket analysis (Harahap & Sulindawaty, 2020). The Apriori algorithm is divided into several stages called narratives. The stages are as follows (Sibarani, 2020):

a. Formation of candidate itemsets. Candidate k-itemsets are formed from combinations of (k-1) itemsets obtained from the previous iteration. The Apriori algorithm method is pruning candidate k-itemsets whose subset contains k-1 items not included in the high-frequency pattern with length k-1.
b. Calculation of support for each candidate k-itemset. The support of each candidate k-itemset is obtained by scanning the database to count the number of transactions that contain all items in the candidate k-itemset. This is also a characteristic of the Apriori algorithm where counting is required by means of the entire database as many as the longest k-itemset.
c. Determine high-frequency patterns. High-frequency patterns containing k-items or k-itemsets are determined from candidate k-itemsets whose support is greater than minimum support.
d. If no new high-frequency patterns are obtained, the entire process is stopped.

## FP-Growth Algorithm

Frequent Pattern Growth (FP-Growth) is an alternative algorithm that can be used to determine the most frequently occurring data set (frequent itemset) in a data collection (Setyo & Wardhana, 2019). The FP-Growth method can be divided into 3 main stages as follows (Fajrin & Maulana, 2018):

a. Conditional pattern base generation stage
   Conditional Pattern Base is a sub-database that contains prefix paths and suffix patterns. The generation of conditional pattern base is obtained through the FP-tree that has been built previously.
b. Conditional FP-Tree generation stage

At this stage, the support count of each item in each conditional pattern base is summed, then each item that has a total support count greater than or equal to the minimum support count will be generated with a conditional FP-tree.

c. Frequent itemset search stage

If the Conditional FP-tree is a single path, then frequent itemsets are obtained by combining items for each conditional FP-tree. If it is not a single path, then FP-Growth generation is performed recursively.

**RapidMiner**

RapidMiner is a powerful software platform for data science and machine learning. It provides a wide range of tools for data preparation, modeling, evaluation, and implementation. RapidMiner is designed to be user-friendly and allows users to easily build and test various models, even without programming experience (Rafi Nahjan et al., 2023).

**Research Flow**

The stages carried out in this study can be illustrated in the figure below:

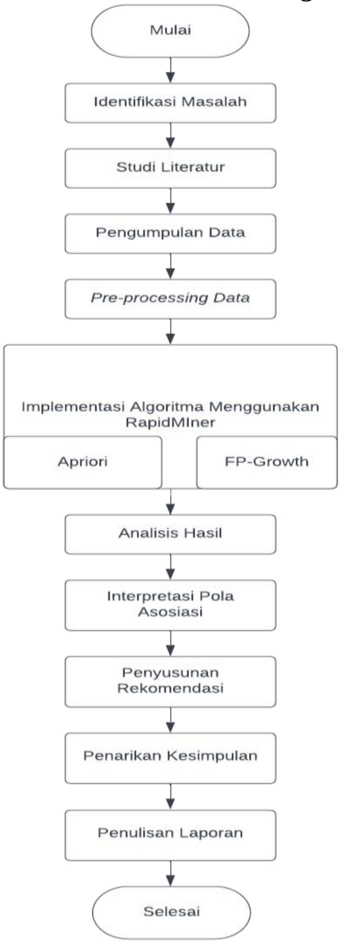


**Figure 1.** Image of Research Flow

## RESULT AND DISCUSSION

**Data Source**

This study uses raw transaction data from the UNWAHA Multi-Purpose Cooperative, for the period of January 1 - December 31, 2023, totaling 31212 records and consisting of 7 variables as shown in the following table.

**Table 1.** Raw Data of Cooperative UNWAHA Sales Transactions

| No | Date | Product | Quantity | Selfling Prices | Purchase Price | Total | Profit |
|---|---|---|---|---|---|---|---|
| 1 | 01/01/2023 | AICE MOCHI CHOCOLATE | 1.00 | 2,365.00 | 3,000.00 | 3,000.00 | 635.00 |
| 2 | 01/01/2023 | CHITATO SPICY MAX CHILI 55G | 1.00 | 6,500.00 | 7,500.00 | 7,500.00 | 1,000.00 |
| 3 | 01/01/2023 | DEKA WHITE JUMBO 16 GR | 1.00 | 875.00 | 1,000.00 | 1,000.00 | 125.00 |
| 4 | 01/01/2023 | GERY SEREAL COKLAT 30G | 2.00 | 1,750.00 | 2,000.00 | 4,000.00 | 500.00 |
| 5 | 01/01/2023 | MIE ENAK 17G | 2.00 | 850.00 | 1,000.00 | 2,000.00 | 300.00 |
| 6 | 01/01/2023 | SIIP COKLAT 5G | 1.00 | 450.00 | 500.00 | 500.00 | 50.00 |
| 7 | 01/01/2023 | SOSIS SO NICE | 2.00 | 800.00 | 1,000.00 | 2,000.00 | 400.00 |
| 8 | 01/01/2023 | TISU NICE 200 SHEETS | 1.00 | 9,500.00 | 15,000.00 | 15,000.00 | 5,500.00 |
| 9 | 01/01/2023 | ULTRA SARI KACANG IJO | 1.00 | 4,900.00 | 5,000.00 | 5,000.00 | 100.00 |
| 10 | 02/01/2023 | ADEM SEJUK CINCAU 350ML | 3.00 | 4,900.00 | 5,000.00 | 15,000.00 | 300.00 |
| 11 | 02/01/2023 | AICE BINGO COOKIES CONE | 4.00 | 4,280.00 | 5,500.00 | 22,000.00 | 4,880.00 |
| 12 | 02/01/2023 | AICE BROWN SUGAR BOBA | 1.00 | 4,850.00 | 6,000.00 | 6,000.00 | 1,150.00 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 31210 | 31/12/2023 | TUJUH KURMA | 1.00 | 9,900.00 | 10,000.00 | 10,000.00 | 100.00 |
| 31211 | 31/12/2023 | ULTRAMILK COKLAT/MOCCA 250ML | 2.00 | 7,400.00 | 7,500.00 | 15,000.00 | 200.00 |
| 31212 | 31/12/2023 | WALENS CHOCO SOES 100G | 1.00 | 8,000.00 | 10,000.00 | 10,000.00 | 2,000.00 |
| | | **Grand Total** | **96,055.00** | | | **485,857,970.00** | **49,958,046.00** |

## Data Preprocessing

In this study, two preprocessing techniques were carried out, namely data preparation and data transformation.

- Data Preparation

  The stage of selecting attributes that will be used for rule formation using the association method which will later be processed through RapidMiner. Only three attributes will be used, including product, date, and quantity as can be seen in the following table:

**Table 2.** Data after attribute selection process

| Date | Product | Quantity |
|---|---|---|
| 01/01/2023 | AICE MOCHI CHOCOLATE | 1.00 |
| 01/01/2023 | CHITATO SPICY MAX CHILI 55G | 1.00 |
| 01/01/2023 | DEKA WHITE JUMBO 16 GR | 1.00 |
| 01/01/2023 | GERY SEREAL COKLAT 30G | 2.00 |
| 01/01/2023 | MIE ENAK 17G | 2.00 |
| 01/01/2023 | SIIP COKLAT 5G | 1.00 |
| 01/01/2023 | SOSIS SO NICE | 2.00 |
| 01/01/2023 | TISU NICE 200 SHEETS | 1.00 |
| 01/01/2023 | ULTRA SARI KACANG IJO | 1.00 |
| 02/01/2023 | ADEM SEJUK CINCAU 350ML | 3.00 |
| 02/01/2023 | AICE BINGO COOKIES CONE | 4.00 |
| 02/01/2023 | AICE BROWN SUGAR BOBA | 1.00 |
| ... | ... | ... |

| ... | ... | ... |
|---|---|---|
| 31/12/2023 | TUJUH KURMA | 1.00 |
| 31/12/2023 | ULTRAMILK COKLAT/MOCCA 250ML | 2.00 |
| 31/12/2023 | WALENS CHOCO SOES 100G | 1.00 |

- Data Transformation
  At this stage, there are three attributes that will undergo data transformation, namely product, date, and quantity. To determine the number of purchases based on product names and also avoid duplication of the same item names, data mapping and grouping from transaction data were carried out as well as a replace missing value process as shown in the table below:

**Table 3.** Data after data transformation process

| Tanggal | 1 DUS CLUB\CLEO GELASAN | ABC EXTRA PEDAS BTL 335ML | ABC KACANG IJO | ABC KACANG IJO 200 ML | ABC KOPI SUSU | ... | ZWITSAL BABY SOAP 70G | ZYLUC STICK |
|---|---|---|---|---|---|---|---|---|
| 01-01-2023 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 |
| 02-01-2023 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 |
| 03-01-2023 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 |
| 04-01-2023 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 |
| 05-01-2023 | 0 | 0 | 0 | 0 | 1 | ... | 0 | 0 |
| 07-01-2023 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 |
| 08-01-2023 | 0 | 0 | 0 | 0 | 2 | ... | 0 | 0 |
| 09-01-2023 | 0 | 0 | 0 | 0 | 2 | ... | 0 | 0 |
| 10-01-2023 | 0 | 0 | 0 | 0 | 1 | ... | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 30-12-2023 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 |
| 31-12-2023 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 |

**Calculation of Support Value Using Ms. Excel**

To find patterns from transaction data, the first thing to do is to calculate the support value of each item obtained using the following equation.

$$Support\ (A) = \frac{Number\ of\ transactions\ containing\ A}{Total\ transactions}$$

The following items, total items sold, and support values found can be seen in the table below:

**Table 4.** Support value for each item

| Produk | Transaksi per Produk | Nilai Support per Produk |
|---|---|---|
| LE MINERALE | 275 | 0,962 |
| GULA 1KG | 191 | 0,668 |
| KRUPUK 500 | 173 | 0,605 |
| AOKA KEJU | 153 | 0,535 |
| MAKARONI ASEP | 153 | 0,535 |
| YUPI ALL VARIAN 5G | 153 | 0,535 |
| KRUPUK PAK JONO | 152 | 0,531 |
| SOSIS SO NICE | 149 | 0,521 |
| … | … | … |
| … | … | … |
| FLORIDINA ALL VARIAN | 89 | 0,311 |
| AICE MOCHI CHOCOLATE | 88 | 0,308 |
| KRIPIK SINGKONG | 88 | 0,308 |
| ISOPLUS | 87 | 0,304 |
| AICE TARO CRISPY | 86 | 0,301 |
| BEAR BRAND 189 GR | 86 | 0,301 |
| ABC KOPI SUSU | 83 | 0,290 |
| AHH KEJU/COKLAT | 78 | 0,273 |
| … | … | … |
| … | … | … |
| ZINC REFRESHING COOL 340ML | 1 | 0,003 |
| ZUPERR KEJU SUSU 130GR | 1 | 0,003 |

After getting the support value of each item, the next step is to select the minimum support value. The author determines the minimum support value of 0.3, 0.29, and 0.27 used for the W-Apriori test and the minimum support value of 0.6, 0.53, 0.52 for the FP-Growth test.

## Implementation of Apriori and FP-Growth Algorithms in RapidMiner

The results of data preprocessing are used as input for further processing. The input data is converted from numeric to binominal using the Numerical to Binominal operator. The Apriori algorithm is tested with the W-Apriori operator, while the FP-Growth algorithm is tested using the FP-Growth operator and Create Association Rules.

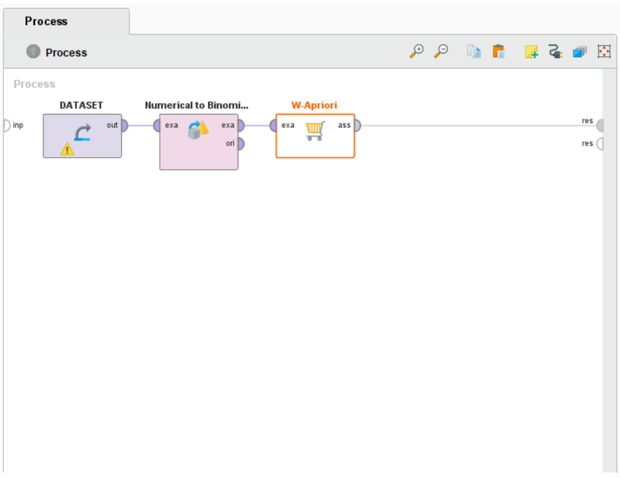


**Figure 2.** Operator Arrangement for W-Apriori Testing

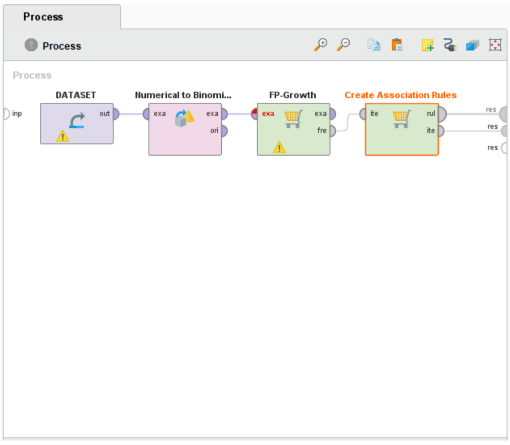


**Figure 3.** Operator Arrangement for FP-Growth Testing

Parameter (C) shows the minimum confidence value, while parameter (M) shows the minimum support value. Next, parameter (N) can be seen, which is the maximum limit of rules to be displayed.

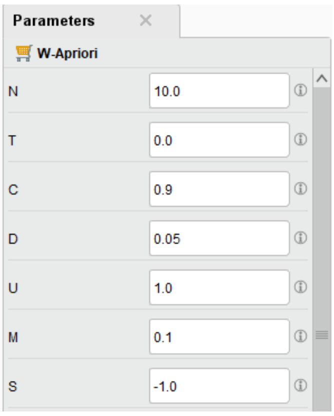


**Figure 4.** W-Apriori Parameters

As for the minimum support and minimum confidence settings of the FP-Growth algorithm, it can be seen in the figure below:

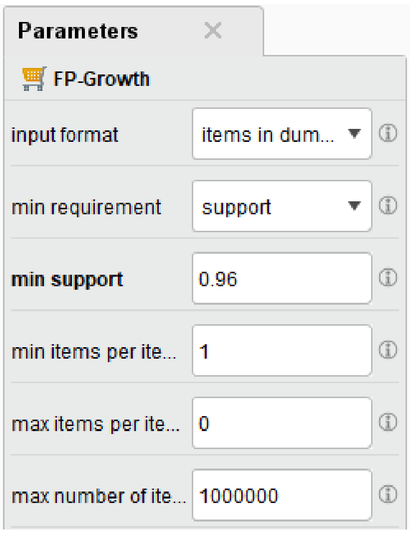


**Figure 5.** FP-Growth Parameters

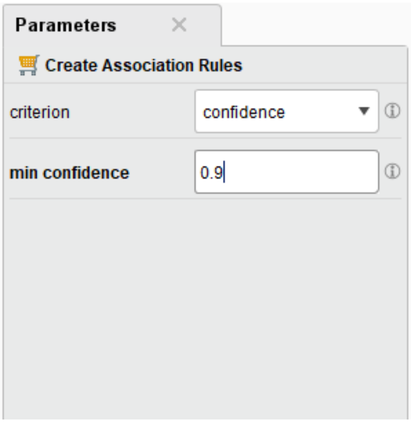


**Figure 6.** Create Association Rules Parameters

The min. support parameter in the FP-Growth operator shows the minimum support value, while the min. confidence parameter in the Create Association Rules operator shows the minimum confidence value.

- Testing using W-Apriori

  From several tests using the Apriori algorithm previously discussed, the researcher concluded that testing with a minimum support value of 0.3 and a minimum confidence value of 0.9 produced 5 rules, which were considered more optimal as they did not generate too many rules that could complicate interpretation, nor too few rules that might miss important patterns. There were also some similarities with the rules produced by testing using the Apriori algorithm.

  The Apriori algorithm produced the highest confidence value of 0.95 or 95% with a combination of 3 items and generated 5 association rules, namely:

  1) If buying MAKARONI ASEP and YUPI ALL VARIAN 5G, then will buy SOSIS SO NICE with a confidence value of 0.95
  2) If buying MAKARONI ASEP and YUPI ALL VARIAN 5G, then will buy KRUPUK PAK JONO with a confidence value of 0.94
  3) If buying KRUPUK 500 and SOSIS SO NICE, then will buy YUPI ALL VARIAN 5G with a confidence value of 0.93
  4) If buying KRUPUK PAK JONO and YUPI ALL VARIAN 5G, then will buy MAKARONI ASEP with a confidence value of 0.91
  5) If buying KRUPUK PAK JONO and YUPI ALL VARIAN 5G, then will buy SOSIS SO NICE with a confidence value of 0.91

- Testing using FP-Growth

  From several tests using the Apriori algorithm previously discussed, the researcher concluded that testing with a minimum support value of 0.52 and a minimum confidence value of 0.9 produced 6 rules, which were considered more optimal as they did not generate too many rules that could complicate interpretation, nor too few rules that might miss important patterns. There were also some similarities with the rules produced by testing using the FP-Growth algorithm.

  The FP-Growth algorithm produced the highest confidence value of 1.00 or 100% with 2 item combinations and generated 6 association rules, namely:

  1) If buying [GULA 1KG], then will buy [LE MINERALE] with a confidence value of 0.974
  2) If buying [YUPI ALL VARIAN 5G], then will buy [LE MINERALE] with a confidence value of 0.987
  3) If buying [KRUPUK 500], then will buy [LE MINERALE] with a confidence value of 0.988
  4) If buying [AOKA KEJU], then will buy [LE MINERALE] with a confidence value of 0.993
  5) If buying [MAKARONI ASEP], then will buy [LE MINERALE] with a confidence value of 1.000
  6) If buying [KRUPUK PAK JONO], then will buy [LE MINERALE] with a confidence value of 1.000

**Discussion**

1) The Apriori and FP-Growth algorithms can be used to discover association rules that reveal relationships between sales products of the UNWAHA Multi-Purpose Cooperative in food and beverage sales.
2) The application of transaction data analysis using the FP-Growth algorithm proved to be faster than the Apriori algorithm with a difference of 11 seconds.
3) The FP-Growth rules testing results proved to be accurate, compared to Apriori which was still less accurate.

**Sales Recommendations for UNWAHA Cooperative**

Based on the analysis of annual sales transaction data from the UNWAHA Cooperative, the researcher can recommend several sales strategies that could be considered for implementation by the Cooperative, based on the rules generated from the application of the Apriori and FP-Growth algorithms. These strategies include the following:

- Marketing Strategy

Implement bundling and discounts for products that are frequently purchased together and are approaching their expiration dates.

- Inventory Management

Arrange product placement in the most easily accessible locations and position them close to each other. For example, food products such as MAKARONI ASEP, KRUPUK PAK JONO, KRUPUK 500, SOSIS SO NICE, and YUPI ALL VARIAN 5G, which are combinations of products often bought together.

And prepare more stock for frequently purchased products. For instance, LE MINERALE, which is the best-selling item.

## CONCLUSIONS

Based on the conducted tests, it can be concluded that both the Apriori and FP-Growth Algorithms can be used to discover rules or association rules that identify relationships between products of the UNWAHA Multi-Purpose Cooperative in sales. There are similarities in some products found in the rules generated by both algorithms, such as MAKARONI ASEP, KRUPUK PAK JONO, and KRUPUK 500. The difference is that the implementation of the FP-Growth algorithm using the RapidMiner tool proved to be faster than the Apriori algorithm by a margin of 11 seconds. The FP-Growth rules testing results proved to be accurate, compared to Apriori which was still less accurate, although it can still be used because its difference from manual calculations is not significant, so its confidence value remains high.

From several tests conducted using different minimum support values, the researcher concluded that testing with a minimum support value of 0.3 and a minimum confidence value of 0.9 for the Apriori algorithm, which produced 5 rules, and testing with a minimum support value of 0.52 and a minimum confidence value of 0.9 for the FP-Growth algorithm, which produced 6 rules, were considered more optimal. This is because they did not generate too many rules that could complicate interpretation, nor too few rules that might miss important patterns, and there were some similarities in the rules produced by both algorithms.

Based on the rules generated by the analysis conducted using annual sales transaction data from the UNWAHA Cooperative, the researcher can suggest several sales strategy plans, such as: implementing bundling and discounts for products that are frequently purchased together and are approaching their expiration dates. Then, arranging product placement in the most easily accessible locations and positioning them close to each other. For example, food products like MAKARONI ASEP, KRUPUK PAK JONO, KRUPUK 500, SOSIS SO NICE, and YUPI ALL VARIAN 5G, which are combinations of products often bought together. And preparing more stock for frequently purchased products. For instance, LE MINERALE, which is the best-selling item.

## REFERENCES

Fajrin, A. A., & Maulana, A. (2018). Penerapan_Data_Mining_Untuk_Analisis_Pol. *Kumpulan Jurnal, Ilmu Komputer*, *05*(01), 27–36.

Harahap, P. N., & Sulindawaty, S. (2020). Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah). *Matics*, *11*(2), 46. https://doi.org/10.18860/mat.v11i2.7821

Rafi Nahjan, M., Nono Heryana, & Apriade Voutama. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 101–104. https://doi.org/10.36040/jati.v7i1.6094

Rerung, R. R. (2018). Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk. *Jurnal Teknologi Rekayasa*, *3*(1), 89. https://doi.org/10.31544/jtera.v3.i1.2018.89-98

Setyo, W. N., & Wardhana, S. (2019). Implementasi Data Mining Pada Penjualan Produk Di Cv Cahaya Setya Menggunakan Algoritma Fp-Growth. *Petir*, *12*(1), 54–63. https://doi.org/10.33322/petir.v12i1.416

Sibarani, A. J. P. (2020). Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *7*(2), 262–276. https://doi.org/10.35957/jatisi.v7i2.195

Widayu, H., Nasution, S. D., Silalahi, N., & Mesran. (2017). Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5. *Media Informatika Budidarma, Vol 1*, *No*(2), 37.