Implementation of Data Mining Using Apriori Algorithm to Determine Inventory of Goods

Ahmed Arifi Hilman Rahman[a1], Zaehol Fatah[a2]

[a]Department of Information Systems, Ibrahimy University, Indonesia
e-mail: [1]ahmedarifihilmanrahman@email.com, [2]zaeholfatah@email.com

Abstrak

Pengusaha yang bergerak di bidang perbelanjaan memiliki prospek yang menjanjikan karena dapat melayani masyarakat kelas bawah dan menengah atas serta memberikan kemudahan bagi masyarakat untuk membeli barang sehari-hari tanpa harus pergi ke supermarket atau pasar swalayan. Namun apabila ketersediaan barang atau bahan yang diperlukan tidak terjamin secara optimal, maka dapat terjadi kekurangan barang atau bahan yang diperlukan. Hal ini juga terjadi di beberapa toko, di mana pelanggan sering kali kehabisan stok berbagai produk dan perlengkapan yang mereka cari, namun hal ini disebabkan oleh kurangnya kebiasaan manajemen inventaris di toko tersebut. Dalam hal ini, tentang mencari tahu produk dan kebutuhan apa yang dibutuhkan pelanggan toko. Dataset ini menggunakan beberapa variabel seperti tanggal transaksi, nama produk, dan jumlah penjualan atau pembelian dengan menerapkan algoritma apriori. Algoritma apriori adalah jenis aturan asosiasi dalam data mining yang digunakan untuk menganalisis dan menemukan pola korelasi. Data yang digunakan dalam penelitian ini adalah sampel 100 data transaksi penjualan. Aturan asosiasi final yang diperoleh dari data transaksi tersebut adalah "Jika konsumen membeli Tepung maka akan membeli Minyak dan Telur" dengan persentase support sebesar 54% dan confidence sebesar 96%. Hasil ini memberikan data nama-nama produk terlaris, yang dapat di gunakan sebagai perkiraan persediaan untuk menghindari kursi kosong yang dapat mengakibatkan kekecewaan pelanggan.

Kata kunci: *Algoritma Apriori, Data Mining, Metode Asosiasi*

Abstract

Entrepreneurs engaged in the shopping sector have promising prospects because they can serve the lower and upper middle classes and provide convenience for people to buy everyday goods without having to go to supermarkets or convenience stores. However, if the availability of goods or materials needed is not optimally guaranteed, there may be a shortage of goods or materials needed. This also happens in some stores, where customers often run out of stock of various products and equipment they are looking for, but this is due to the lack of inventory management habits in the store. In this case, it is about finding out what products and needs are needed by store customers. This dataset uses several variables such as transaction date, product name, and sales or purchase amount by applying the apriori algorithm. The apriori algorithm is a type of association rule in data mining that is used to analyze and find correlation patterns. The data used in this study is a sample of 100 sales transaction data. The final association rule obtained from the transaction data is "If consumers buy Flour, they will buy Oil and Eggs" with a support percentage of 54% and a confidence of 96%. These results provide data on the names of the best-selling products, which can be used as an inventory estimate to avoid empty seats that can result in customer disappointment.

Keywords : *Apriori Algorithm, Data Mining, Association Method*

1. Introduction

Data mining is the process of extracting useful knowledge or interesting patterns from a large dataset. The main goal of data mining is to find previously unseen patterns that can provide valuable insights and support good decision making. Data mining is the analytical

process of finding meaningful relationships, patterns and behaviors in large amounts of data, using statistical, mathematical, artificial intelligence and machine learning techniques to extract useful and relevant information and knowledge from a database.[1] Data mining has seven basic functions, namely Prediction, Sequencing, Classification, Association, Clustering, Forecasting, Description.[2] Data mining also plays a key role in helping organizations to make better decisions, understand customers better, and even predict future trends.[3]

Managing inventory is a significant challenge for businesses. If a store has too much stock compared to consumer demand, the store will suffer losses due to unsold goods, especially for products that have an expiration date or are easily damaged. Conversely, if the store only has a minimum amount of stock that is not sufficient for demand, this can cause customers to leave because the desired item is not available. We often find that consumers want to buy a product but the item is not available, so they switch to another place, and in the future, they may hesitate to shop again at that store because they are worried about running out of goods. Of course, problems like this can reduce the level of sales in the store. To solve this problem, an intelligent data mining system is needed that can help in making decisions regarding the goods and stock to be offered.

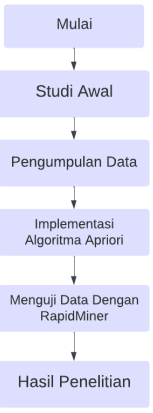
The Apriori algorithm is an algorithm that performs frequent itemset searches using the association rule technique.[4] The Apriori algorithm allows the identification of association patterns in transaction data, allowing us to find relationships and correlations between frequently occurring items. The reliability of the Apriori algorithm is based on strong mathematical principles and has a solid theoretical basis. This provides confidence that the association patterns found have high statistical reliability. The basic concept of the Apriori algorithm is relatively simple and easy to understand, allowing for fairly easy application in data analysis.[5] Therefore, the apriori algorithm is expected to be able to solve problems regarding inventory by associating which goods must be available.

This method will be used to analyze and create a system related to the products that consumers want to buy. It is expected that this will provide benefits to shop owners as a consideration in designing their sales strategies.

2. Research Method / Proposed Method

2.1. Framework

A framework is a process stage that starts from the beginning to the end. A framework is a series of steps that guide in solving a problem. Here is the working structure of the Association Rule method.



2.2. Preliminary Study

At this stage, the researcher's focus is to formulate the background, objectives, and problems to be discussed. Some steps taken at this stage include:.

a) Studying the problem
b) Determining the scope of the problem
c) Studying some literature
d) Data analysis

2.3. Data collection

Data collection is carried out in order to obtain primary data needed for research purposes. While in the study of data mining application, there are two research methods that can be used to collect data on the availability of goods, including:

a) Conducting library research is a method used to use libraries and journal books as sources of information in determining the factors, parameters, and tables that will be used in research.
b) Field work research is a type of research that is carried out directly using various methods, such as reading literature to obtain information, collecting data, studying data, validating data, and looking for references related to the case in the research.

2.4. Data Mining

Data Mining is a process of automatically searching for useful information in large data storage. Data mining techniques are used to examine large databases as a way to find new and useful patterns. However, not all information search work can be stated as data mining. Data mining has many methods, the data mining method used is Knowledge Discovery in Database (KDD) or can be called Pattern Recognition.[6]

2.5. Apriori Algorithm

The apriori algorithm is a data mining analysis method that focuses on searching for frequent itemsets using the Association Rule technique. Another definition of the apriori algorithm is a method for finding relationships between one or more itemsets in a dataset, which is usually widely used in transaction data called market baskets.[7]

2.6. Association Rules

Association rules are one of the methods that aim to find patterns that often appear among many transactions, where each transaction consists of several items so that this method will support the recommendation system through the discovery of patterns between items in the transactions that occur.[8] The basic methodology of association analysis is divided into two stages:

a) High frequency pattern analysis
b) This stage searches for a combination of items that meet the minimum requirements of the support value in the database. The support value of an item is obtained by the following formula:

$$Support(A) = \frac{Jumlah\ Transaksi\ Mengandung\ A}{Total\ Transaksi} X\,100$$

(1)

Meanwhile, the Support value of 2 items is obtained from the following formula

$$Support(A,B) = \frac{Jumlah\ Transaksi\ Mengandung\ A\ dan\ B}{Total\ Transaksi} X\,100$$

(2)

c) Formation of associative rules

After all high frequency patterns have been successfully identified, then an association rule is sought that meets the minimum requirements for confidence by calculating the confidence of the associative rule A U B.

$$Confident = P(A,B) = \frac{Jumlah\ Transaksi\ Nilai\ A\ dan\ B}{Total\ Transaksi} X\,100$$

(3)

2.7. RapidMiner

Rapid Miner is an open source software. Rapid Miner is a solution for analyzing Data Mining, text mining, and predictive analysis. Rapid Miner uses various descriptive and predictive methods or techniques to provide insight to users, so they can make the best decisions.[9]

3. Result and Discussion

3.1. Calculation Stage

Currently, as the amount of data used increases, it takes time to complete the mapping, so the author completed the survey in June 2017 using sample data from commodity transaction data. Convert commodity trading data to Excel data format and enter it into the Apriori algorithm formula to get a shortcut.

To design the Apriori Algorithm processing in goods transactions to determine inventory, the following data is required:

a) Item sample data

Table 1Item sample data

No	*Name of goods*
1	Eggs
2	Oil
3	Flour
4	Soy Sauce
5	Milk
6	Snack
7	Candy
8	Detergent
9	Soap

b) Transaction data sample

Table 2 Transaction data sample

Transaction Code	*transaction date*	*Name of goods*
1	*01/06/2017*	Eggs, Oil, Candy, Flour
2	*01/06/2017*	Detergent, Soy Sauce, Milk, Soap
3	*01/06/2017*	Eggs, Oil, Soy Sauce, Snack, Candy, Flour
4	*02/06/2017*	Detergent, Soy Sauce, Soap
5	*02/06/2017*	Eggs, Oil, Soy Sauce, Snack, Candy, Flour
6	*02/06/2017*	Eggs, Detergent, Oil, Soap, Snack, Candy, Flour
7	*03/06/2017*	Eggs, Detergent, Oil, Soap, Snack, Candy, Flour
8	*03/06/2017*	Eggs, Detergent, Oil, Soy Sauce, Soap, Snack, Candy
9	*03/06/2017*	Eggs, Oil, Soy Sauce, Milk, Snack, Candy, Flour
10	*03/06/2017*	Eggs, Detergent, Oil, Soy Sauce, Milk, Flour
…	…	…
99	*30/06/2017*	Detergent, Soy Sauce, Milk, Candy, Flour
100	*30/06/2017*	Soy Sauce, Snack, Flour

The steps for calculating data using the Apriori Algorithm are as follows:

1) Transaction Tabular Data

This stage functions as data that is input into RapidMiner by first creating the data in tabular form, which can be seen in Table 3.

Table 3 Transaction tabular data

transaction date	Eggs	Oil	Flour	Soy Sauce	Milk	Snack	Candy	Detergent	Soap
1	1	1	1	0	0	0	1	0	0
2	0	0	0	1	1	0	0	1	1
3	1	1	1	1	0	1	1	0	0
4	0	0	0	1	0	0	0	1	1
5	1	1	1	1	0	1	1	0	0

6	1	1	1	0	0	1	1	1	1
7	1	1	1	0	0	1	1	1	0
8	1	0	0	1	0	1	1	0	1
9	1	1	1	1	1	1	1	0	0
10	1	1	1	1	1	0	0	1	0
…	…	…	…	…	…	…	…	…	…
99	1	0	0	0	0	1	0	1	0
100	0	0	1	1	0	1	0	0	0

2) Data 1 Itemset

Based on the available transaction data, the initial step is to find 2 candidate itemsets with the following support values:

Table 4 Data 1 Itemset

No	***Name of goods***	***Quantity***	***Support***
1	Eggs	*81*	*0.81*
2	Oil	*27*	*0.72*
3	Flour	*67*	*0.67*
4	Soy Sauce	*57*	*0.57*
5	Milk	*55*	*0.55*
6	Snack	*52*	*0.52*
7	Candy	*51*	*0.51*
8	Detergent	*51*	*0.51*
9	Soap	*50*	*0.50*

Based on table 4 which contains items with support values, by setting a minimum support of 5%, items that meet these requirements can be found in table 5.

Table 5 Support Value Results 1 Itemset with Minimum

No	***Name of goods***	***Quantity***	***Support***	****100%***
1	Eggs	*81*	*0.81*	*81*
2	Oil	*72*	*0.72*	*72*
3	Flour	*67*	*0.67*	*67*
4	Soy Sauce	*57*	*0.57*	*57*
5	Milk	*55*	*0.55*	*55*
6	Snack	*52*	*0.52*	*52*
7	Candy	*51*	*0.51*	*51*
8	Detergent	*51*	*0.51*	*51*
9	Soap	*50*	*0.50*	*50*

3) Data 2 Itemset

Looking for 2 candidate itemsets with the following support values:

Table 6 Data 2 Itemset

No	***Name of goods***	***Quantity***	***Support***
1	Egg, Flour	*61*	*0.61*
2	Egg, Oil	*64*	*0.64*
3	Egg, Snack	*45*	*0.45*
4	Egg, Candy	*45*	*0.45*
5	Egg, Detergent	*41*	*0.41*
6	Egg, Soy Sauce	*41*	*0.41*
7	Egg, Milk	*41*	*0.41*

8	Egg, Soap	*40*	*0.40*
9	Flour, Oil	*56*	*0.56*
10	Flour, Snack	*45*	*0.45*
11	Flour, Candy	*41*	*0.41*
12	Flour, Detergent	*36*	*0.36*
13	Flour, Soy Sauce	*38*	*0.38*
14	Flour, Milk	*36*	*0.36*
15	Flour, Soap	*39*	*0.39*
16	Oil, Snack	*37*	*0.37*
17	Oil, Candy	*38*	*0.38*
18	Oil, Detergent	*33*	*0.33*
19	Oil, Soy Sauce	*35*	*0.35*
20	Oil, Milk	*32*	*0.32*
21	Oil, Soap	*35*	*0.35*
22	Snack, Candy	*32*	*0.32*
23	Snack, Detergent	*34*	*0.34*
24	Snack, Soy Sauce	*34*	*0.34*
25	Snack, Milk	*26*	*0.26*
26	Snack, Soap	*26*	*0.26*
27	Candy, Detergent	*28*	*0.28*
28	Candy, Soy Sauce	*28*	*0.28*
29	Candy, Milk	*32*	*0.32*
30	Candy, Soap	*26*	*0.26*
31	Detergent, Soy Sauce	*29*	*0.29*
32	Detergent, Milk	*28*	*0.28*
33	Detergent, Soap	*26*	*0.26*
34	Soy Sauce, Milk	*29*	*0.29*
35	Soy Sauce, Soap	*26*	*0.26*
36	Milk, Soap	*27*	*0.27*

Based on table 6, where the items have been sorted based on their support value, by setting a minimum support value of 5%, the items that meet the requirements can be found in table 7.

Table 7 Support Value Results 2 Itemset with Minimum

No	***Name of goods***	***Quantity***	***Support***	****100%***
1	Eggs, Flour	*61*	*0.61*	*61*
2	Eggs, Oil	*64*	*0.64*	*64*
3	Flour, Oil	*56*	*0.56*	*56*

4) Data 3 Itemset

Looking for 3 Itemset candidates with the following support values:

Table 8 Data 3 Itemset

No	***Name of goods***	***Quantity***	***Support***
1	Eggs, Flour, Oil	*54*	*0.54*
2	Eggs, Flour, Snacks	*38*	*0.38*
3	Eggs, Flour, Candy	*35*	*0.35*
4	Eggs, Flour, Detergent	*32*	*0.32*
5	Eggs, Flour, Soy Sauce	*33*	*0.33*
6	Eggs, Flour, Milk	*32*	*0.32*
7	Eggs, Flour, Soap	*33*	*0.33*
8	Eggs, Oil, Snacks	*36*	*0.36*
9	Eggs, Oil, Candy	*36*	*0.36*
10	Eggs, Oil, Detergent	*32*	*0.32*
11	Eggs, Oil Soy Sauce	*34*	*0.34*

12	Eggs, Oil, Milk	*30*	*0.30*
13	Eggs, Oil, Soap	*34*	*0.34*
14	Eggs, Snacks, Candy	*26*	*0.26*
15	Eggs, Snacks, Detergent	*27*	*0.27*
16	Eggs, Snacks, Soy Sauce	*29*	*0.29*
17	Eggs, Snacks, Milk	*20*	*0.20*
18	Eggs, Snacks, Soap	*21*	*0.21*
19	Eggs, Candy, Detergent	*24*	*0.24*
20	Eggs, Candy, Soy Sauce	*23*	*0.23*
21	Eggs, Candy, Milk	*26*	*0.26*
22	Eggs, Candy, Soap	*22*	*0.22*
23	Eggs, Detergent, Soy Sauce	*23*	*0.23*
24	Eggs, Detergent, Milk	*21*	*0.21*
25	Eggs, Detergent, Soy Sauce, Milk	*20*	*0.20*
26	Eggs, Soy Sauce, Soap	*22*	*0.22*
27	Eggs, Milk, Soap	*20*	*0.20*
28	Flour, Oil, Snack	*22*	*0.22*
29	Flour, Oil, Candy	*34*	*0.34*
30	Flour, Oil, Detergent	*32*	*0.32*
31	Flour, Oil, Soy Sauce	*28*	*0.28*
32	Flour, Oil, Milk	*31*	*0.31*
33	Flour, Oil, Soap	*27*	*0.27*
34	Eggs, Flour, Oil	*31*	*0.31*
35	Flour, Snack, Candy	*25*	*0.25*
36	Flour, Snack, Detergent	*26*	*0.26*
37	Flour, Snack, Soy Sauce	*26*	*0.26*
38	Flour, Snack, Milk	*20*	*0.20*
39	Flour, Snack, Soap	*20*	*0.20*
40	Flour, Candy, Detergent	*20*	*0.20*
41	Flour, Candy, Soy Sauce	*23*	*0.23*
42	Flour, Candy, Milk	*24*	*0.24*
42	Flour, Candy, Soap	*22*	*0.22*
44	Flour, Detergent, Soy Sauce	*20*	*0.20*
45	Flour, Soy Sauce, Milk	*22*	*0.22*
46	Flour, Milk, Soap	*20*	*0.20*
47	Oil, Snack, Candy	*22*	*0.22*
48	Oil, Snack, Detergent	*21*	*0.21*
49	Oil, Snack, Soy Sauce	*24*	*0.24*
50	Oil, Candy, Detergent	*20*	*0.20*
51	Oil, Candy, Soy Sauce	*20*	*0.20*
52	Oil, Candy, Milk	*21*	*0.21*
53	Snack, Candy, Soy Sauce	*20*	*0.20*
54	Snack, Detergent, Soy Sauce	*21*	*0.21*
55	Snack, Soy Sauce, Milk	*20*	*0.20*

Based on the data in table 8 which includes items and their support values, by setting a minimum support of 5%, items that meet the minimum requirements can be found in table 9.

Table 9 Support Value Results 3 Itemset with Minimum

No	***Name of goods***	***Quantity***	***Support***	****100%***
1	*Eggs, Flour, Oil*	*54*	*0.54*	*54*

5) Formation of Association Rules

To find the relationship of the rules to the previous steps, the next step is to calculate the confidence value for each item listed in table 8 using the formula mentioned earlier. Then, the values for Support and Confidence from the

combination of 3 itemset patterns with the minimum support requirement = 5% and minimum confidence = 8% have been listed in table 10.

Table 10 Results of Support and Confidence calculations

No	Name of goods	Support	Cnfidence
1	Flour, Oil, Eggs	*0.54 %*	*0.96 %*
2	Eggs, Flour, Oil	*0.54 %*	*0.88 %*
3	Eggs, Oil, Flour	*0.54 %*	*0.84 %*

Based on the process that has been gone through, the items that achieve the relevant Support value and Confidence results are in accordance with the association rules that have been formed, then the following conclusions can be drawn:

a) If consumers buy flour, they will buy oil and eggs with support of 0.54% and confidence of 0.96% and the store will provide these goods.
b) If consumers buy eggs, they will buy flour and oil with support of 0.54% and confidence of 0.88% and the store will provide these goods.
c) c) If consumers buy eggs, they will buy oil and flour with support of 0.54% and confidence of 0.84% and the store will provide these goods.

3.2. Implementation On RapidMiner

a) Import Data to Repository

At this stage, the data to be processed is prepared, in the form of all sales transaction data that has been selected.

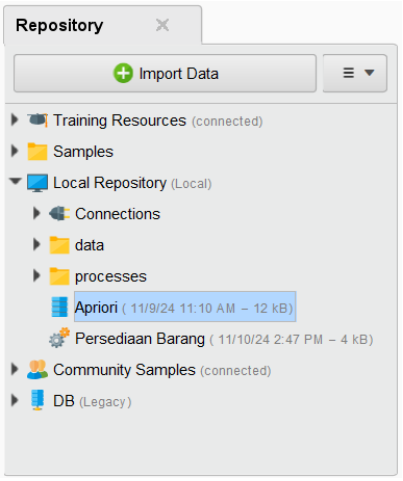


Figure 1 Import Data to Repository

b) Operator Design

This step involves dragging and dropping data tables into the workflow. Numeric operators will be converted to binomial form to modify the attribute values used. Next, add the fp-growth operator and the create Association Rule operator by setting a Minimum Support of 0.5 on the fp-growth operator and a Minimum Confidence of 0.8 on the create Association Rule operator. Once all operators are composed, connect all existing operators.

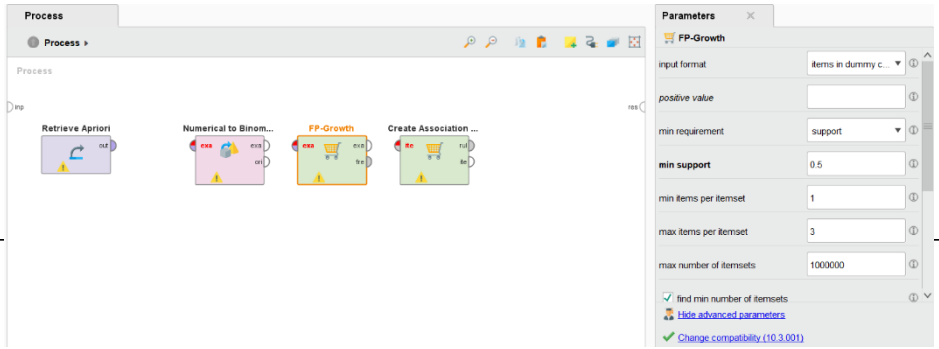


Figure 2 Minimum Support Fill

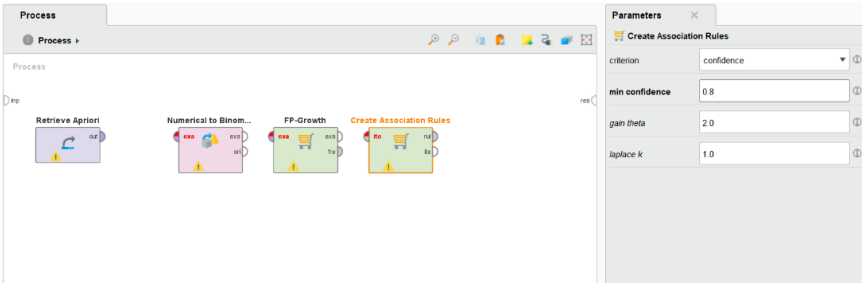


Figure 3 Minimum Confidence Filling

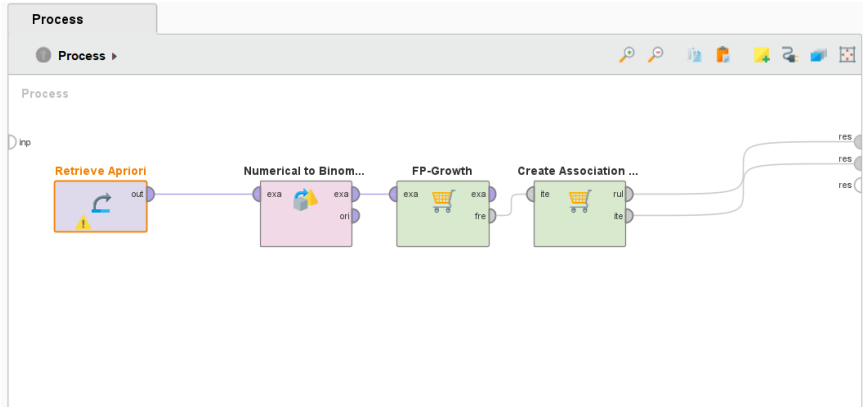
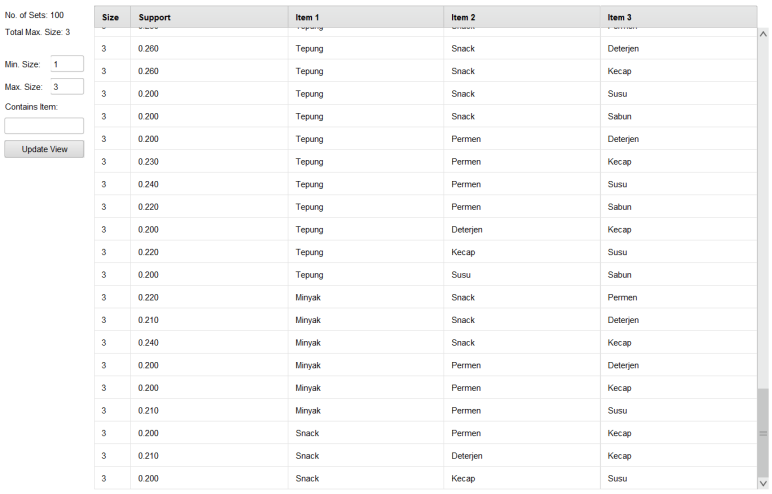


Figure 4 Operator Design

c) Frequent Itemset and Association Rule Results

This step is the final stage in data mining using rapid miner. At this stage, three Frequent Itemsets and Association Rules were found that meet the Minimum Support and Minimum Trust criteria. The results can be seen in Figure 6 for Frequent Itemsets and Figure 7 for Association Rules formed in rapidminer.



Size	Support	Item 1	Item 2	Item 3
3	0.260	Tepung	Snack	Deterjen
3	0.260	Tepung	Snack	Kecap
3	0.200	Tepung	Snack	Susu
3	0.200	Tepung	Snack	Sabun
3	0.200	Tepung	Permen	Deterjen
3	0.230	Tepung	Permen	Kecap
3	0.240	Tepung	Permen	Susu
3	0.220	Tepung	Permen	Sabun
3	0.200	Tepung	Deterjen	Kecap
3	0.220	Tepung	Kecap	Susu
3	0.200	Tepung	Susu	Sabun
3	0.220	Minyak	Snack	Permen
3	0.210	Minyak	Snack	Deterjen
3	0.240	Minyak	Snack	Kecap
3	0.200	Minyak	Permen	Deterjen
3	0.200	Minyak	Permen	Kecap
3	0.210	Minyak	Permen	Susu
3	0.200	Snack	Permen	Kecap
3	0.210	Snack	Deterjen	Kecap
3	0.200	Snack	Kecap	Susu

Figure 5 Frequent Item Sets

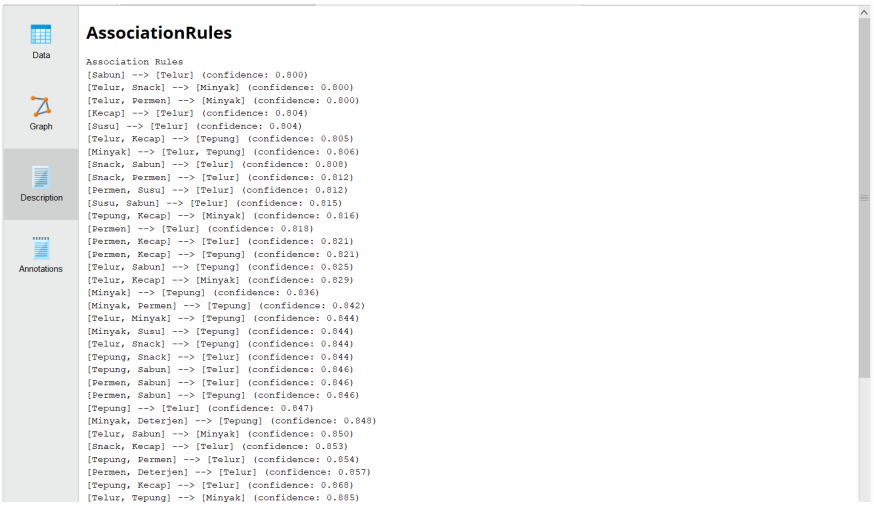


Figure 6 Association Rules

4. Conclusion

Based on the results and discussions that have been carried out, it can be concluded that the implementation of data mining using the Apriori algorithm can be used to predict the results of staple food sales in stores and find out what items should be sold. Items will be available if they are available. The larger the collection of elements contained in each piece of data, the more relationships there are.

Based on the research results, it is known that the highest demand is for the purchase of wheat flour and the purchase of oil and eggs, with a confidence value of 96%. The final result is the same if calculated in Microsoft Excel using the RapidMiner application and using the Apriori algorithm.

References

[1] C. Carudin *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: https://books.google.co.id/books?id=m-QGEQAAQBAJ

[2] Y. Ardilla *et al.*, *DATA MINING DAN APLIKASINYA*. Penerbit Widina, 2021. [Online]. Available: https://books.google.co.id/books?id=53FXEAAAQBAJ

[3] S. K. M. M. S. I. Ir. T. Irfan Fajri *et al.*, *Data Mining*. Serasi Media Teknologi, 2024. [Online]. Available: https://books.google.co.id/books?id=YykdEQAAQBAJ

[4] Z. MUSIAFA, *Algoritma Apriori Penentuan Pola Penjualan: Studi Kasus Toko Akhtar Galaxy*. ZAYID MUSIAFA, 2021. [Online]. Available: https://books.google.co.id/books?id=3CFQEAAAQBAJ

[5] R. F. Putra *et al.*, *DATA MINING : Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023. [Online]. Available: https://books.google.co.id/books?id=zLHGEAAAQBAJ

[6] M. S. Iskandar and Z. Fatah, "Gudang Jurnal Multidisiplin Ilmu Implementasi Metode Algoritma K-Means Clustering Untuk Menentukan Penerima Program Indonesia Pintar (PIP)," vol. 2, no. November, pp. 1–8, 2024.

[7] Muhammad Alwi, Ninis Anggraini, and Rodia, "Analisis Data Mining Pada Pemilihan Jenis Game Terpopuler Menggunakan Algoritma Apriori," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 11, no. 1, pp. 9–15, 2023, doi: 10.21063/jtif.2023.v11.1.9-15.

[8] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data

Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang," *J. Janitra Inform. dan Sist. Inf.*, vol. 2, no. 1, pp. 9–19, 2022, doi: 10.25008/janitra.v2i1.142.

[9] Z. Fatah and P. Diagnosis, "Penerapan Data Mining Untuk Prediksi Diagnosis Demam Berdarah Dengan Algoritma Decision Tree," vol. 3, no. May, pp. 63–71, 2025.