http://ejournal.seaninstitute.or.id/index.php/InfoSains
Jurnal Info Sains : Informatikan dan Sains, Volume 13, No 01 Maret 2023
E-ISSN.2797-7889, P-ISSN.2089-3329

Apriori Algorithm Testing Using The Rapidminer Application

[1]Paska Marto Hasugian, [2]Pandi Barita Nauli Simangunsung
[1,2]Program Studi Rekayasa Perangkat Lunak, STMIK Pelita Nusantara, Politeknik LP3I Medan
Email: **Paskamarto86@gmail.com**

Keywords

Apriori
Rapidminer
aplication

Abstract. In this study, the Apriori algorithm was tested using the RapidMiner application. Apriori algorithm is one of the algorithms used in association analysis in data mining. The purpose of this research is to test the performance of Apriori algorithm in identifying association patterns in the given dataset. The research method used is experimentation using relevant datasets. Testing is done by implementing the Apriori algorithm in the RapidMiner application, which provides a visual interface and various functions for data analysis. During testing, adjustments were made to Apriori algorithm parameters, such as support and confidence, to optimize analysis results. The data generated from the tests were evaluated to determine significant association patterns. The test results show the success rate of the Apriori algorithm in identifying association patterns in the given dataset. The performance evaluation of the algorithm is done based on relevant metrics, such as accuracy, execution speed, and association pattern quality score. The conclusion of this research is that testing the Apriori algorithm using the RapidMiner application can provide adequate results in association analysis. This research has important implications in the field of data mining and association analysis, especially in the use of the Apriori algorithm using the RapidMiner application. It is hoped that this research can contribute to the development of association analysis techniques and improve understanding of the performance of the Apriori algorithm in the context of its use in practical applications.

1. INTRODUCTION

Data mining is the process of extracting valuable information from large and complex datasets. One of the techniques in data mining is association analysis, which aims to identify patterns of association or strong relationships between items or attributes in a dataset. (Kusumo et al., 2019). For example, in a business context, association analysis can be used to identify patterns of product purchases that often occur together. lThe Apriori algorithm is one of the popular algorithms used in association analysis. It works based on association rules, which look for association patterns that meet certain support and confidence criteria. (Mawengkang & Budhiarti Nababan, 2015). For example, if a number of items co-occur in a dataset with significant frequency and high confidence, then it is considered a significant association pattern.

Testing the Apriori algorithm is important in the context of association analysis. Testing aims to evaluate the performance of the algorithm in finding significant association patterns, as well as to identify the strengths and weaknesses of the algorithm. In previous research, several factors such as dataset size, algorithm parameters, and the influence of noise on the data have been studied in testing the Apriori algorithm, RapidMiner is a popular software in the field of data mining and predictive analysis. It provides various tools and functions that make it easy for users to perform data analysis, including Apriori algorithm testing. RapidMiner offers an intuitive visual interface, the ability to import data from various formats, and data pre-processing features that support efficient Apriori algorithm testing. (Marto Hasugian, 2016).

Limitations of Previous Research: Although several previous studies have tested the Apriori algorithm, there are still shortcomings and aspects that have not been covered. For example, some studies only focus on relatively small datasets or do not consider the complexity of algorithm parameters that can affect the performance of the Apriori algorithm. Therefore, testing Apriori algorithm using RapidMiner application is still relevant and necessary to complete and enrich our knowledge about the performance of this algorithm in the context of more complex association analysis. (Listriani et al., n.d.).

Dewi Listriani et.al's research, they applied the Apriori algorithm to analyze transaction data or purchases made by consumers at Gramedia Bintaro Bookstore. Association methods, such as those used by the Apriori algorithm, can reveal patterns of purchases that occur simultaneously or are

related among items purchased by consumers. By applying this method to a consumer shopping pattern analysis application, the research or project may aim to identify relationships between certain products or items that are often purchased together by consumers at Gramedia Bintaro Bookstore. The results of this analysis can provide valuable insights for the bookstore, for example, to devise more effective marketing strategies, the arrangement of goods in the store, or even the development of new products. (Listriani et al., n.d.).

Several studies have applied the Apriori algorithm to find the relationship between the type of imported commodity and the amount of monthly demand. This research can provide a better understanding of market demand patterns related to the type of imported commodity. By identifying the relationship between the type of imported commodity and the amount of monthly demand, this research can help businesses or stakeholders in understanding consumer preferences and market demand related to certain imported products. By identifying the relationship between import commodity types and monthly demand, this research can provide important information for stock planning and management. For example, the results of the research can assist in determining the optimal inventory levels for each type of imported commodity, avoiding potentially costly shortages or overstocks. (*602-Article Text-2726-1-10-20211217*, n.d.)

Research by Rizal Sidiq Al Amin Hasil et.al in research that has tested data experiments using manual calculations and Rapid Miner software using the Apriori Algorithm which aims to determine the inventory of goods. (*300969-Penerapan-Algoritma-Apriori-Untuk-Menemu-D7a1ce63*, n.d.), Dina Nurhidayanti's research 2022 The results of research on finding association rules with the apriori algorithm through manual calculations show the same results as the RapidMiner application, namely 3 association rules with the highest Confiedence value of 75% resulting in rules: if you buy Piece Set Slide then buy Spark Plugs, if you buy Rear Lights then buy Front Brake Pads, and if you buy Front Lights then buy Front Brake Pads. Based on these rules, the a priori algorithm is very useful for knowing the frequent-itemset pattern of spare parts that sell best and the association rules or relationships between each product that is purchased simultaneously. (Nurhidayanti et al., 2022).

Ina Maryani's research, 2022 concluded that the results of calculating data mining information using the apriori algorithm, sales transaction information at the GOC Kosmetik store with a minimum support limit of 60% and a minimum confidance of 70%. The most product sales are when buying SS to wanting to buy KN with a support value of 69% and a confidance value of 88% so that this data can provide advice to the store to improve products and further improve product sales strategies to achieve a large profit value. (Maryani et al., 2022).

2. METHOD

Apriori algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 to determine Frequent itemsets for Boolean association rules. Apriori algorithm is a type of association rule in data mining. Rules that express associations between several attributes are often called affinity analysis or market basket analysis Association analysis or association rule mining is a data mining technique for finding rules for a combination of items. One of the stages of association analysis that attracts the attention of many researchers to produce efficient algorithms is high-frequency pattern analysis (frequent pattern mining).

3. RESULTS AND DISCUSSION

3.1 Data Set Preparation Stage

The research process, the author uses goods transaction data as material to complete the research taken from Toko X. The goods transaction data will be converted into excel 2013 data format then the data is entered into the Apriori Algorithm formula to get the Association In designing the processing of the Apriori Algorithm on goods transactions in determining the inventory of goods, data is required as Sample Data for Basic Food Items sold at Store X can be seen from Table 1.Tabel 1. Data sample item barang sembako

Item	Name
1	Oil
2	Sugar
3	Tea
4	Chicken Eggs
5	Flour
6	Rice
7	Salt
8	Coffee
9	Milk
10	Instan Noodles

with sample sales transaction data at shop x can be seen in table 2:

Table 2. Sales transaction data d shop x

Item	Name
1	Rice, Salt, Eggs Chicken, Oil, Sugar
2	Coffee, Sugar, Tea, Milk, Oil, Rice
3	Chicken Eggs, Ground Pepper, Soy Sauce, Rice, Salt, Noodles, Instant Pot
4	Oil, sugar, cigarettes, lighters, milk, rice
5	Salt, Chicken Egg, Flour, Oil
6	Chicken Eggs, Salt, Rice, Cigarettes
7	Lighters, Cigarettes, Rice, Oil, Sugar, Flour
8	Salt, Shampoo, Detergent, Flour, Chicken Egg, Oil
9	Oil, Sugar, Salt, Flour, Rice
10	Sugar, Salt, Chicken Eggs, Oil, Rice

3.2 Implementation

Below are the working steps for implementing the rapid miner, which are as follows:

1. Tabular Format

Tabular format is the input data used to perform the Apriori testing process with rapid miner, in this test used is excel which is stored with Tabular values with entries 0 and 1 with the meaning that 1 is an available value and o is not available.

Table 3 1. Data Preparation in Excel

Oil	Sugar	Tea	Egg	Flour	Rice	Salt	Coffe	Milk	Instant Noodles
1	1	1	1	0	1	1	0	0	0
1	1	0	0	0	1	0	1	1	0
0	0	0	1	0	1	1	0	0	1
1	1	0	0	0	1	0	0	1	0
1	0	0	1	1	0	1	0	0	0
0	0	0	1	0	1	1	0	0	0
1	1	0	0	1	1	0	0	0	0
1	0	0	1	1	0	1	0	0	0
1	1	0	0	1	1	1	0	0	0
1	1	0	1	0	1	1	0	0	0

2. Import data In Rapid Miner

Data import is done using the "Read CSV" or "Read Excel" operator, depending on the format of the file to be imported. These operators allow users to insert data from CSV or Excel files into the workflow. To start importing data, users need to add the "Read CSV" or "Read Excel" operator to the work canvas. Once the operator is added, users can click on it to configure the data import settings. In the operator settings, users can select the data file to be imported by using the "Browse" or "Choose File" button. Once the file is selected, users can set other settings such as delimiter for CSV files or sheet for Excel files. Users can also select specific columns that they want to import. Once the settings are completed, users can click the "Apply" or "OK" button to confirm the settings. The data import operator will connect to the next operator in the work process, so that the data can be used and processed according to the analysis needs.

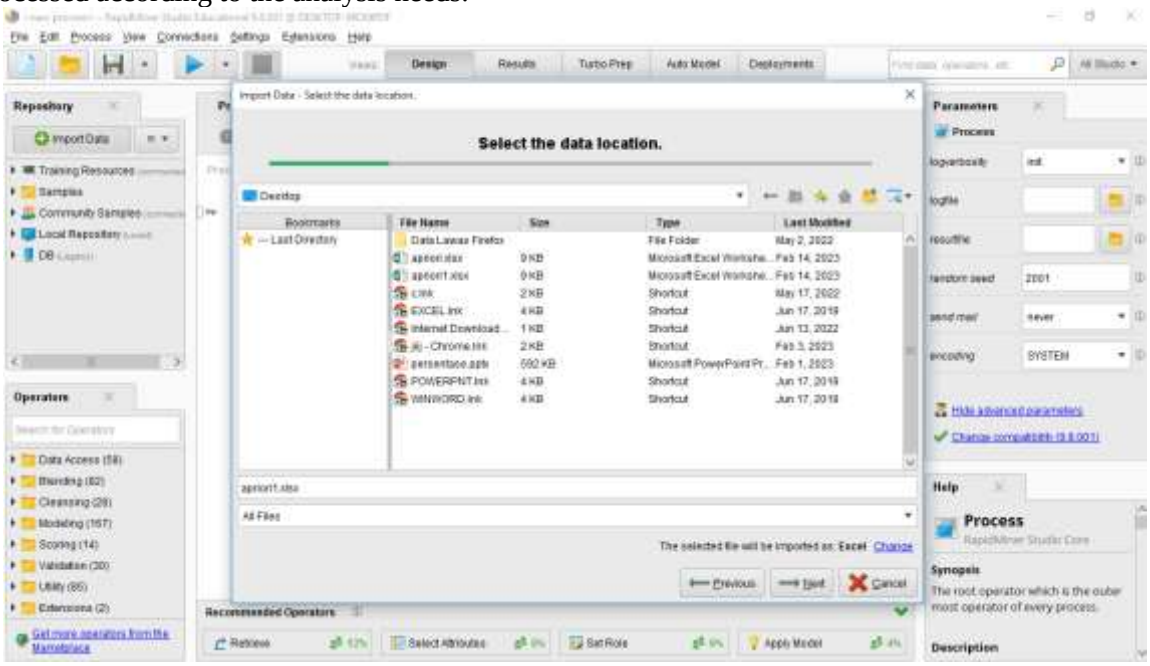

Figure 2. Import data Set

With the selection in accordance with the predetermined data set, the select the cells to import data will appear, which is explained in the following figure

http://ejournal.seaninstitute.or.id/index.php/InfoSains
Jurnal Info Sains : Informatikan dan Sains, Volume 13, No 01 Maret 2023
E-ISSN.2797-7889, P-ISSN.2089-3329

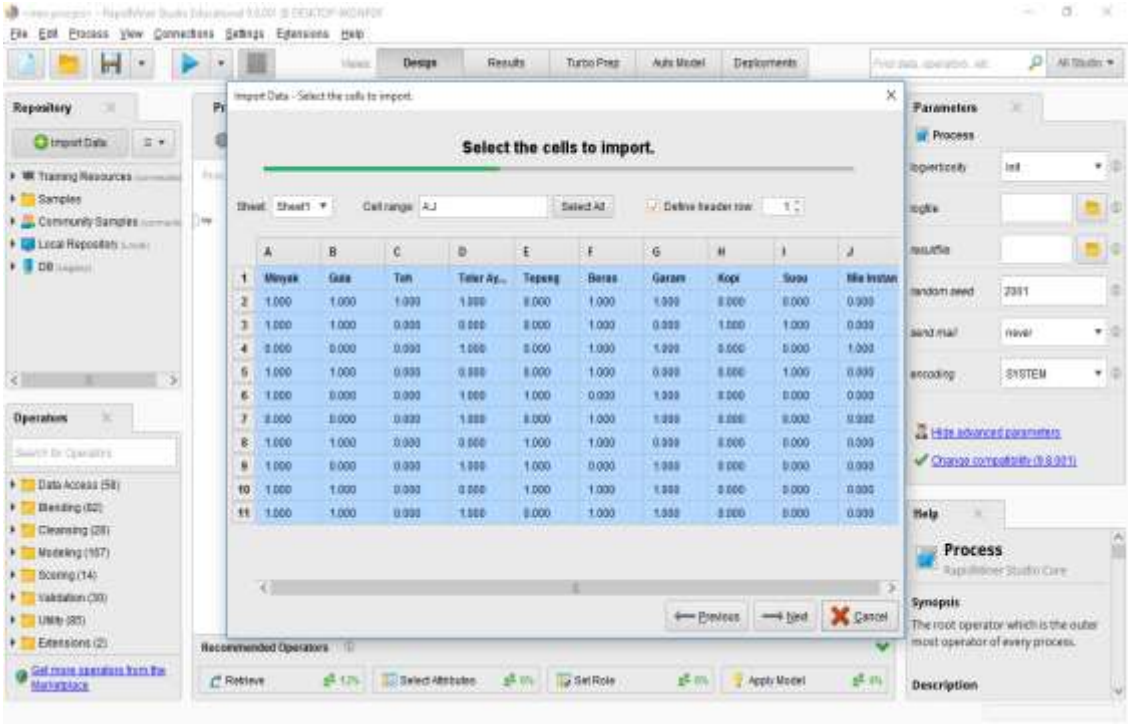

Figure 3. Cell to Import

After the cell to import stage and the Next selection is made so that the data set appears in the rapid miner application, described in the following figure:

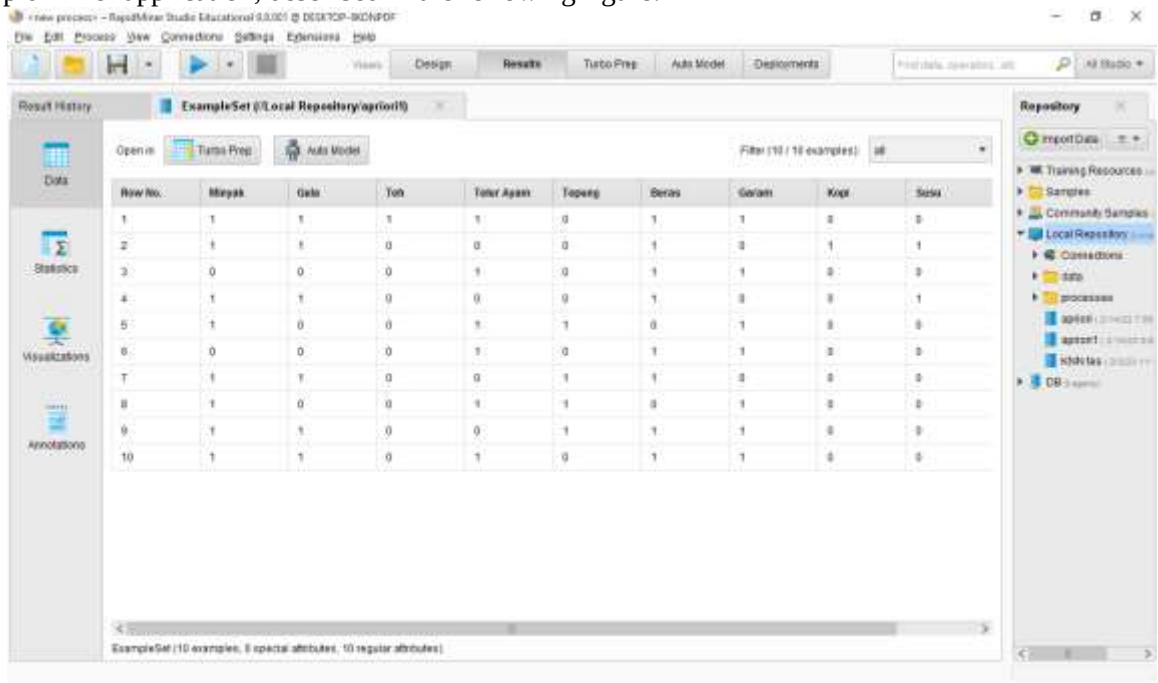

Figure 4. Data Set View

Then to the Design section we Drag the Apriori then input Discritive Frequency, Numerical Binomal then don't forget to enter Fp Growt and Create Ascotiation then draw a line like the image below

http://ejournal.seaninstitute.or.id/index.php/InfoSains
Jurnal Info Sains : Informatikan dan Sains, Volume 13, No 01 Maret 2023
E-ISSN.2797-7889, P-ISSN.2089-3329

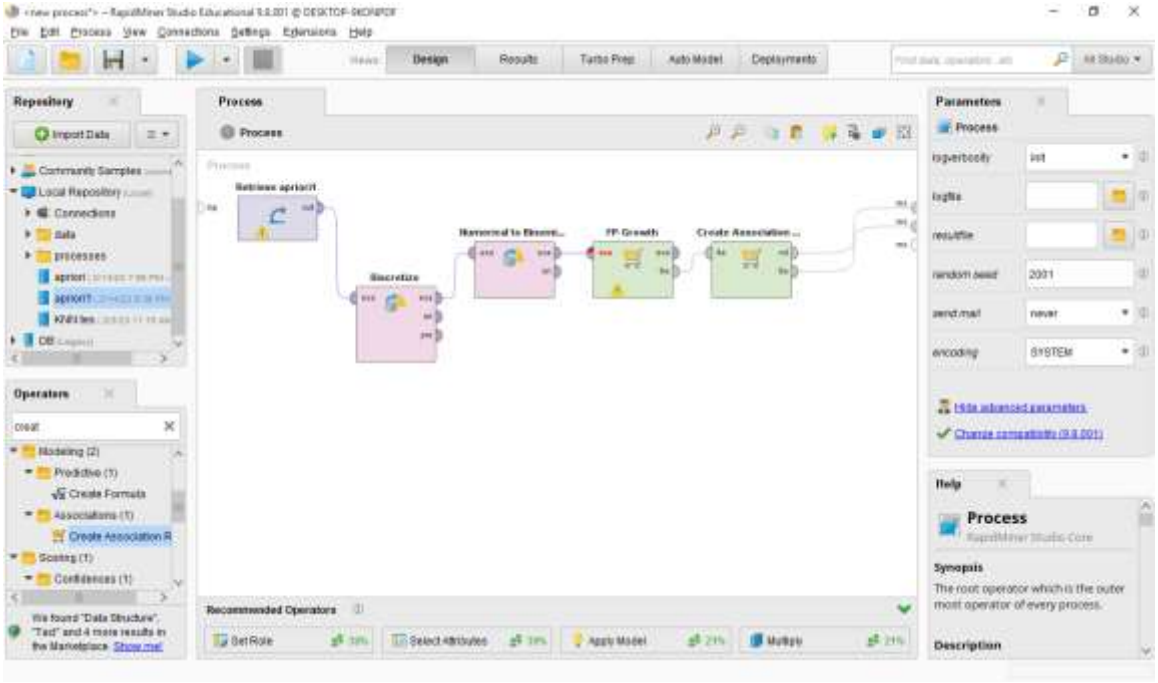

Figure 5. Design Structure Rapid Miner

After the architecture is in accordance with the working principles of apriori, the next step is to run by selecting run so that the analysis data will appear as shown in the following figure:

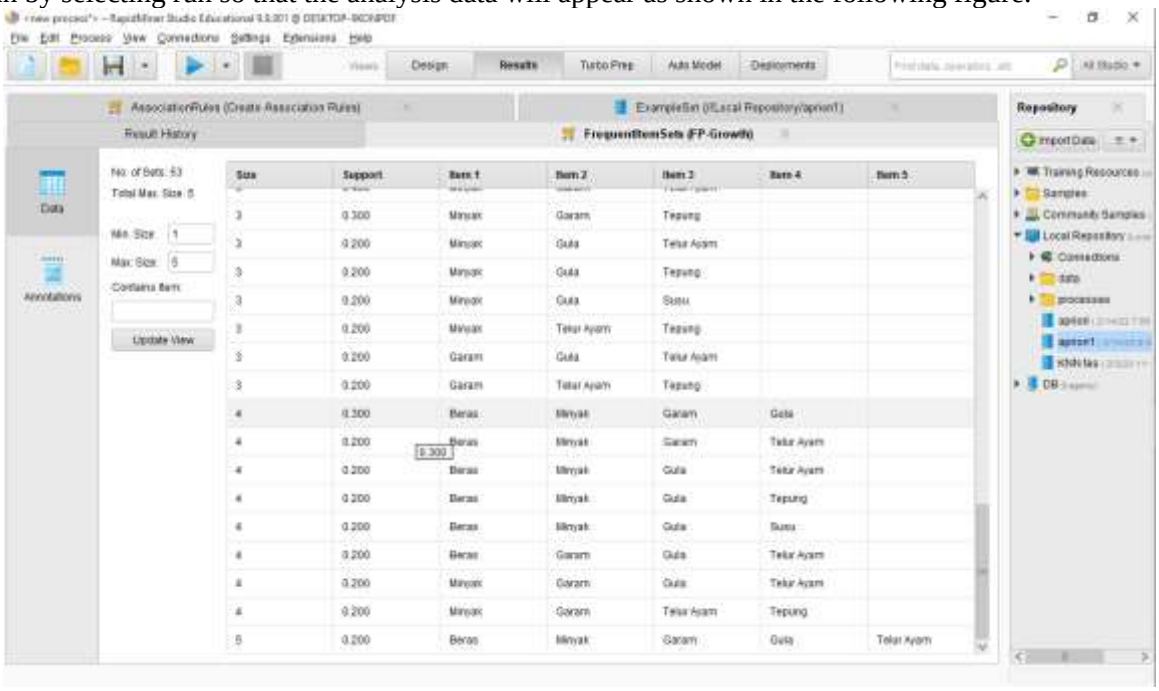

Figure 6. Analysis Result

In the picture above, there is a table view that shows the results of sales transactions that have met the minimum support and minimum confidence requirements. By choosing to see the Asscotiation bias as shown below

http://ejournal.seaninstitute.or.id/index.php/InfoSains
Jurnal Info Sains : Informatikan dan Sains, Volume 13, No 01 Maret 2023
E-ISSN.2797-7889, P-ISSN.2089-3329

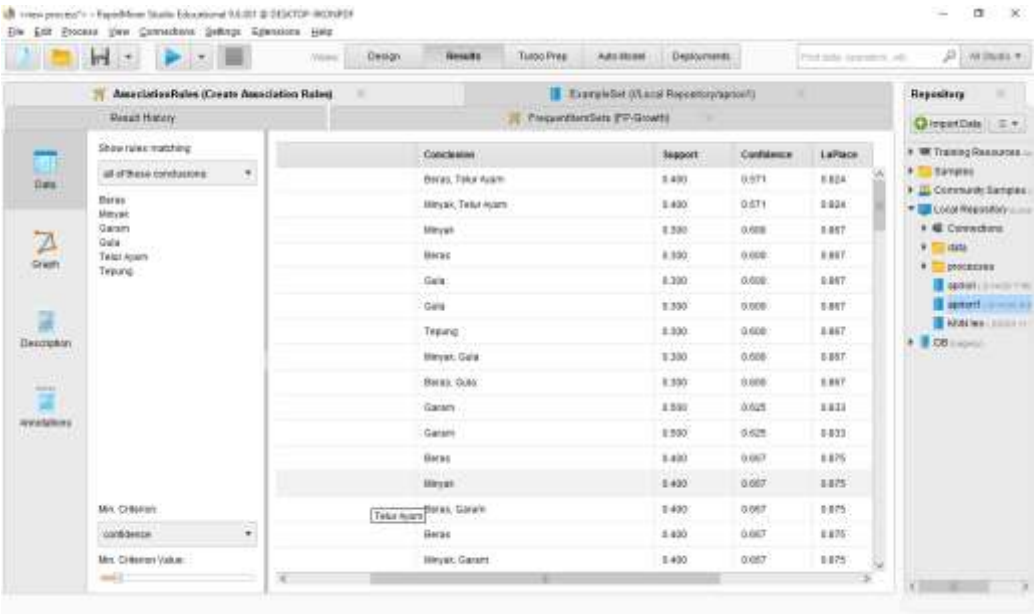

Figure 7. Confidence Result

To display the connectivity of each rule shown in the following display:

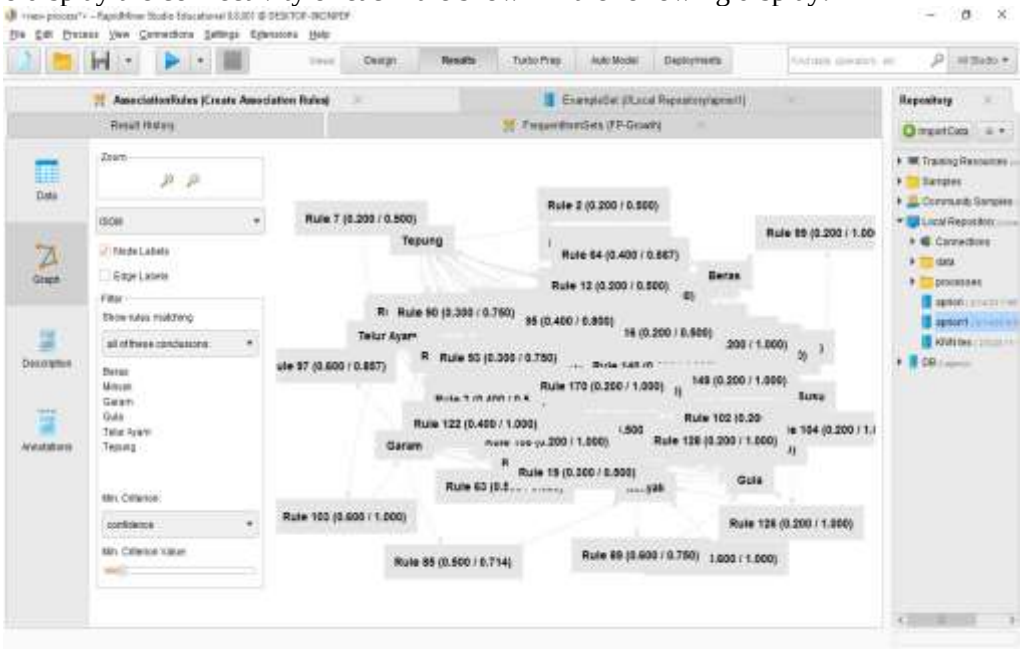

Figure 8. Graph of Connectedness between Rules

4. CONCLUSION

Based on the results and discussions that have been carried out, it can be concluded that the application of data mining using the Apriori Algorithm can be applied in predicting the sales results of basic food items at Toko X in order to find out what items should be in stock can be applied. The more the number of itemsets in each data, the more relationships there will be in each data. that testing the Apriori algorithm using the RapidMiner application has advantages in performance, efficiency, and ease of use. This application can be a useful tool in association analysis and assist users in identifying relevant association patterns in their datasets.

REFERENCE

602-Article Text-2726-1-10-20211217. (n.d.).

300969-penerapan-algoritma-apriori-untuk-menemu-d7a1ce63. (n.d.).

Kusumo, H., Sediyono, E., & Marwata, M. (2019). Analisis Algoritma Apriori untuk Mendukung Strategi Promosi Perguruan Tinggi. *Walisongo Journal of Information Technology*, *1*(1), 49. https://doi.org/10.21580/wjit.2019.1.1.4000

Listriani, D., Setyaningrum, A. H., & Eka, F. (n.d.). *PENERAPAN METODE ASOSIASI MENGGUNAKAN ALGORITMA APRIORI PADA APLIKASI ANALISA POLA BELANJA KONSUMEN (Studi Kasus Toko Buku Gramedia Bintaro)*. *9*(2).

Marto Hasugian, P. (2016). *Junal Manajemen dan Informatika Komputer Pelita Nusantara*. *19*(1).

Maryani, I., Revianti, O., & Muhammad Nur, H. (2022). Implementasi Data Mining Pada Penjualan Di Toko GOC Kosmetik Dengan Menggunakan Metode Algoritma Apriori. In *Indonesian Journal on Software Engineering (IJSE)* (Vol. 8, Issue 1). http://ejournal.bsi.ac.id/ejurnal/index.php/ijse92

Mawengkang, H., & Budhiarti Nababan, E. (2015). *PENGEMBANGAN ALGORTIMA APRIORI UNTUK PENGAMBILAN KEPUTUSAN THE DEVELOPMENT APRIORI ALGORITHM FOR DECISION-MAKING* (Vol. 4, Issue 2).

Nurhidayanti, D., Kurniawati, I., & Artikel, S. (2022). Implementasi Algoritma Apriori Dalam Menemukan Association Rules Pada Persediaan Sparepart Motor INFORMASI ARTIKEL ABSTRACT. *INNOVATION IN RESEARCH OF INFORMATICS*, *4*(2), 62–67.