

# Apriori Algorithm Analysis to Determine Purchasing Patterns at Beleven Farma Pharmacy

Fara Lufiah<sup>1</sup>, Dwi Rosa Indah<sup>2</sup>, Mgs. Afriyan Firdaus<sup>3</sup>

<sup>1,2,3</sup> Faculty of Computer Science, Sriwijaya University

[09031182126015@student.unsri.ac.id](mailto:09031182126015@student.unsri.ac.id)<sup>1</sup>, [indah812@unsri.ac.id](mailto:indah812@unsri.ac.id)<sup>2</sup>, [mafriyan\\_firdaus@unsri.ac.id](mailto:mafriyan_firdaus@unsri.ac.id)<sup>3</sup>

## Article Info

### Article history:

Received 2024-11-12

Revised 2024-12-03

Accepted 2025-01-18

### Keyword:

*Apriori Algorithm,  
Purchasing Patterns,  
Transaction Data Analysis,  
Product Bundling,  
Knowledge Discovery in  
Database (KDD).*

## ABSTRACT

Beleven Farma Pharmacy is a place that provides medicines and other health products such as supplements, vitamins and also various health tests. As a newly established pharmacy, no innovations have been made to improve sales strategies. Analysis of purchasing patterns can produce information that helps pharmacies in determining product bundling recommendations as well as determining product layout. This research applies the a priori algorithm method and uses rapidminer tools to identify drug purchasing patterns from transaction data at the Beleven Farma pharmacy. The Knowledge discovery in database (KDD) method is used as a reference in the data processing process. Based on tests carried out by the author, the resulting rules are that if you buy hemaviton you will buy vice with 4% support and 91% confidence and if you buy amoxicillin you will buy paracetamol with 4% support and 64% confidence. Thus, the resulting information can be used to support decision making in determining marketing strategies so as to increase sales at pharmacies.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. INTRODUCTION

In the contemporary business environment, which is characterised by heightened competition, a comprehensive grasp of consumer purchasing patterns is a crucial determinant of business success. Effective data management enables organisations to identify patterns, comprehend consumer behaviour and enhance operational efficiency. The healthcare sector, too, is confronted with the challenge of competition in the provision of medical services[1]. Effective drug inventory management can ensure adequate availability, prevent stock shortages that can hamper services, and optimize clinic expenditures[2]. The high consumer demand for medicine, along with the increase in the number of pharmacists, is a driving factor in the increase in the number of pharmacies. This situation has indirectly led to intensified competition among pharmacies to attract as many customers as possible[3]. Beleven Farma Pharmacy is one of the pharmacies located in the city of Palembang, this pharmacy provides a variety of medicinal needs and also provides diabetes, cholesterol and uric acid testing services.

Based on interviews with the pharmacy manager as a newly established business, sales still tend to be low because this pharmacy does not yet have a deep understanding of customer preferences or market trends that can be used to formulate effective strategies. In addition, the layout of products within the pharmacy is still not well organized. Products may not be placed in strategic locations, resulting in many medicines being out of sight or difficult for customers to access. Therefore, proper analysis is needed to increase sales and competitiveness of pharmacies amid the dynamics of the growing business world. One of them is by utilizing the field of data mining. Data mining is a way to find interesting patterns and information in selected data using certain methods, then it will produce knowledge[4]. This research uses the Knowledge Discovery in Database (KDD) approach. Knowledge Discovery in Database (KDD) is the process of identifying more valuable, understandable, and new information from large and complex data sets[5]. This research utilizes the apriori algorithm method. Apriori algorithm is one of the efficient data mining algorithms to identify association patterns in transaction data. This algorithm works by finding items that are often purchased

together in purchase transactions, as well as assisting in establishing association rules among those items[6]. This helps to provide information on product placement, pricing, promotions, and the products most frequently purchased by consumers [7].

Research related to data mining analysis using the apriori algorithm that can be used as a reference includes research conducted by Nugraheni & Nugroho on market basket analysis with the apriori algorithm, based on this research, analysis using the apriori algorithm can optimize the highest item set combination pattern to increase sales by placing products close together and offering attractive promotions based on the combination of items generated from the research, the goal is to maintain customer satisfaction on an ongoing basis[8]. In addition, research conducted by Hilmam discusses analyzing sales data at deliberative grocery stores using the apriori algorithm. Based on this research using the a priori algorithm can reveal significant purchasing patterns. The research also identifies products with the lowest sales and provides information so that stores can optimize inventory strategies and product arrangements, thereby increasing sales efficiency[9].

**II. METHOD**

The research method consists of the stages of work that will be carried out by researchers, here is the flow of the research stages :

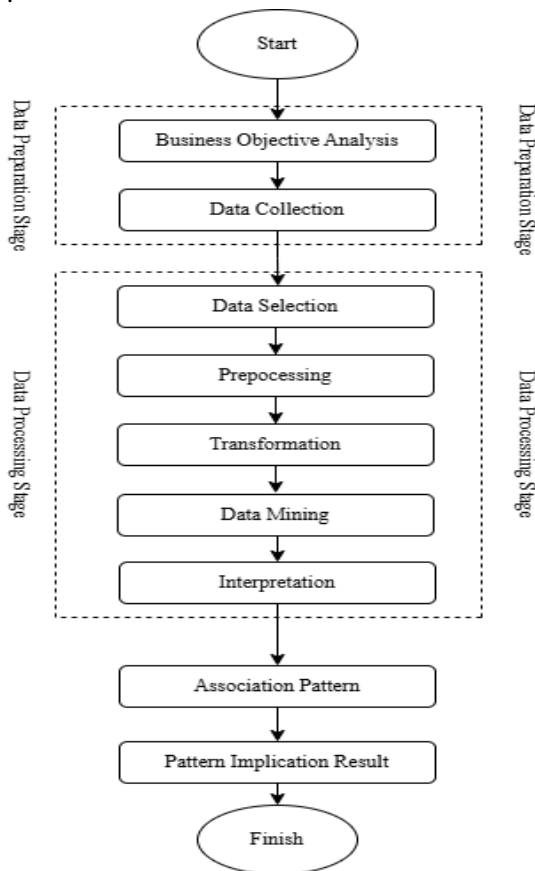


Figure 1. Flow of Research Stages

**A. Business Objective Analysis**

Processing purchase transaction data at the beleven farma pharmacy by analyzing data mining techniques using the a priori algorithm is carried out to find out what strategies can be applied. The results of the analysis are patterns related to purchases that commonly occur at beleven farma pharmacies, this information allows pharmacies to determine product bundling and organize product placement layouts so that they can assist in improving sales strategies.

**B. Data Collection**

*1) Primary Data*

- Interviews, conducted to the owner of the beleven farma pharmacy, this method is used in the data collection process to find problems that must be researched and extract information related to sales transaction data for the beleven farma pharmacy.
- Observation, in this study, the authors directly observed the sales process that occurred at the beleven farma pharmacy.
- Documents, archival data on sales transactions at the beleven farma pharmacy during the period of August and September 2024.

*2) Secondary Data*

Data sekunder pada penelitian ini yaitu studi dokumen. Metode ini dilakukan dengan cara mengumpulkan dan menganalisis berbagai dokumen yang relevan dengan topik. Secondary data in this research is a document study. This method is carried out by collecting and analyzing various documents relevant to the research topic.

**C. Data Processing**

The data analysis process in this study uses the Knowledge discovery in database (KDD) method. Knowledge discovery in database (KDD) is a data analysis method commonly used in research related to data mining[10].

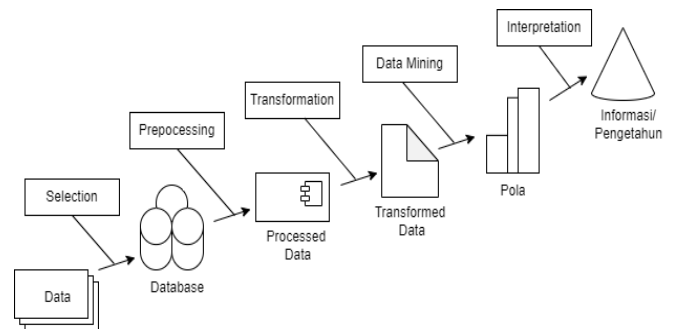


Figure 2. Stages of Knowledge Discovery in Database

The steps of the KDD process consist of 5 stages including; selection, namely the selection of relevant data, preprocessing in the form of cleaning data from missing and unnecessary data, transformation is changing data into an appropriate form, data mining to find patterns or information on selected data using certain methods or techniques, and evaluation of interesting patterns in the form of certain representations [11].

The selection of the Knowledge Discovery in Databases (KDD) method is based on its ability to use data mining techniques to find useful and valuable information from data. This process involves the application of algorithms to identify patterns contained in data, so as to uncover significant insights and support better decision making [12]. The following describes the steps in the Knowledge Discovery in Databases (KDD) process.

1) *Data Selection*

At the data selection stage, the data required in the research is selected, namely sales transaction data at the Beleven Farma pharmacy. This stage is needed so that the data processing process carried out can be in accordance with the objectives of the study. The data source (dataset) used is sales transaction data for the August and September 2024 periods totaling 907 data. The product categories analyzed were items that were sold more than 4 times, this was done because of the limited number of transactions. If all products are included without restriction, it will be difficult to find a significant relationship.

TABLE 1. TRANSACTION DATA

Transaction No.	Produk Name	Transaction. date	Total Qty	Total Amount
1	teosal	01-08-2024	2	7000
2	le minerale	01-08-2024	9	5000
	antangan permen herbal			
3	piroxicam trifa 10mg	01-08-2024	20	5000
4	paramex	01-08-2024	1	3000
5	simvastatin hexpharm 10 mg	01-08-2024	1	4000
6	sanmol syr 120 ml	01-08-2024	1	21500
7	Gentamicin 0.1% cream 5gr	01-08-2024	1	5000
8	tolak angin cair plus madu	01-08-2024	3	22000
	vitamin d3 1000 ipi ecer (racikan)			
9	hufagrip bp syr 60 ml	01-08-2024	1	21000
10	antangan permen herbal	01-08-2024	9	15000
	curcuma force			
...	...	...	...	...
907	bodrex extra	31-09-2024	1	3000

Based on all the attributes in the transaction data table, only two types of fields will be used for the testing process, namely:

- Product Name = This attribute contains information about the product code in the sales transaction data table.
- Transaction No = This attribute contains information about the sales transaction number contained in the transaction data table.

2) *Preprocessing*

Preprocessing is also known as data cleaning, which aims to remove inconsistent, duplicate, or empty data[13]. Based on this explanation, it can be seen that the data cleaning

process is a step taken to remove noise and inconsistent or irrelevant data.

TABLE 2. PREPROCESSING RESULT DATA

Transaction No.	Produk Name
1	teosal
2	le minerale btl 600 ml
	antangan permen herbal
3	piroxicam trifa 10mg
4	paramex
5	simvastatin hexpharm 10 mg
6	sanmol syr 120 ml
7	gentamicin kimia farma 0.1% cream 5gr
8	tolak angin cair plus madu 15 ml
	vitamin d3 1000 ipi ecer (racikan)
9	hufagrip bp syr 60 ml
10	antangan permen herbal
	curcuma force
...	...
906	bodrex extra
	amoxicillin hexpharm 500mg
907	bodrex extra

At this stage, the August and September sales transaction data will be merged. The combined August and September sales transaction data that has gone through the selection stage will be cleaned so that it is ready for analysis. This process includes deleting incomplete or duplicate data, correcting entry errors, and converting data formats for consistency.

TABLE 3. TABULAR DATA

No.	Vitacimin 500 Mg	Hemaviton Stamina Plus	Vicee 500	Salonpas Koyo	...	Mefenamic Acid Triman 500mg
1	0	0	0	0	...	0
2	0	0	0	0	...	0
3	0	0	0	0	...	0
4	0	0	0	0	...	0
5	0	0	0	0	...	0
...	...	...	...	...	...	...
14	0	0	0	1	...	0
15	1	0	0	0	...	0
16	0	0	0	0	...	0
17	0	0	0	0	...	0
18	0	0	0	0	...	0
19	0	0	1	0	...	0
20	0	0	0	0	...	0
21	0	0	0	0	...	0
22	0	1	0	0	...	0
...	...	...	...	...	...	...
907	0	0	0	0	...	0

3) *Transformation*

Transformation aims to transform data into a format suitable for analysis, this process involves summarization or aggregation operations to facilitate data mining[14]. In the tabular table, the colom section contains 101 product names and the row section contains 907 transaction numbers.

This process aims to prepare data in a structured format, making it easier to apply the a priori algorithm in data mining analysis in Rapidminer. Rapidminer is open source software built with the Java programming language, so it can be used on various operating systems. Rapidminer provides solutions for data mining analysis by applying descriptive and predictive techniques, allowing users to make better decisions [15]. Rapidminer can analyze text, extract patterns from datasets, and integrate these results with statistical methods, databases, and artificial intelligence, thus generating valuable information from data processing [16].

#### 4) Data Mining

Data mining is a data analysis technique based on statistical applications and aims to extract information [17]. At this stage, the analysis is carried out by utilizing data mining algorithms, namely the association method using the a priori algorithm. The reason for choosing the apriori algorithm as the method to be used in analyzing this problem is because the apriori algorithm is an algorithm that is useful for applying association rules in determining the frequency of itemsets. This algorithm can identify relationships between one or more items in a dataset. Apriori is very suitable for grouping items that are often purchased together at one time, managing product inventory, and determining product purchase patterns[18]. The following is the calculation process in the Apriori algorithm[19]:

- *1-itemset formation:*

This process is done by identifying the highest frequency patterns in the transaction data. Technically, it involves searching for items that meet the minimum support criteria. The author has the freedom to set the minimum support and minimum confidence values based on the need. The author can specify, if the goal is to find data with strong association relationships, can set high minimum support and minimum confidence values. However, if we want to see the variation of data without paying too much attention to how strong the association relationship is between items, we can use lower minimum support and minimum confidence values.

- *2-itemset formation:*

In this stage, we look for candidate 2-itemsets that are a combination of items that have met the minimum support value requirement in the previous stage (1-itemset).

- *3-itemset formation:*

The next stage is to search for 3-itemset candidates. This process includes the combination of items that have met the minimum support value requirement from the formation of 2-itemset.

- *Formation of Association Rules:*

The next step is to build association rules by confirming if the minimum confidence value is met. This is done by calculating the Confidence value for each association rule, where the statement reads “if A, then B”.

#### 5) Interpretation

This stage is part of the Knowledge Discovery in Databases (KDD) process which includes checking whether the patterns or information found are relevant to the research objectives. Information patterns generated from the data mining process need to be presented in a format that is easily understood by interested parties[20]. At this stage, the lift ratio will be calculated to measure the validity of the resulting association rules.

#### D. Association Pattern

The result of the data processing that has been done is in the form of association rules. In these rules there are patterns or relationships between data that have been processed in the previous stages. These association patterns will be identified in order to generate knowledge/information that can be used in the process of making better decisions based on customer behavior and preferences, thereby improving efficiency, business strategy, marketing, and product management.

#### E. Implication Pattern Results

After knowing the association pattern that occurs between one item and another, the pattern will be analyzed for its implications for the business. This stage aims to understand customer preferences in order to determine appropriate strategies to increase sales.

### III. RESULTS AND DISCUSSION

In the ensuing discourse, a series of calculations will be conducted with a view to establishing relationships between discrete items. This will be achieved through the utilisation of the Apriori algorithmic approach.

#### A. Itemset Formation

##### 1) 1-Itemset Formation

This stage aims to identify individual items that meet predefined minimum criteria based on the Support values in the database. This process involves a thorough analysis of each item in the dataset to determine how often each item appears in recorded transactions. Here is an example of calculating the support value:

$$\begin{aligned} \text{Support } A &= \frac{\text{Jumlah Transaksi Mengandung } A}{\text{Total Transaksi}} \times 100\% \dots (1) \\ \text{Support (Vitacimin)} &= \frac{\text{Jumlah Vitacimin}}{\text{Total Transaksi}} \times 100\% \\ \text{Support Vitacimin} &= \frac{29}{907} \times 100\% \\ \text{Support Vitacimin} &= 3,1\% \end{aligned}$$

The results of the Support 1-itemset calculation can be seen in the table 4. There are a total of 101 products for which the support value is calculated. The results of the calculation of the support value will then be selected which ones meet the minimum support value set, namely by setting a minimum support value of 4%. The minimum support value applied in

this analysis is already at the maximum possible level. If this value is increased further, it is likely that there will be no detectable relationship between products, given the very large number of products and transactions that are still limited.

TABLE 4.  
SUPPORT CALCULATION RESULT DATA

Product Name	Total Transaction	Support
vitacimin 500 mg lemon	29	3,1%
vicee 500	75	8,2%
salonpas koyo	7	0,7%
samcofenac 50mg	9	0,9%
scopma plus	9	0,9%
onemed test kehamilan instant	13	1,4%
ipi vitamin c jeruk	16	1,7%
yusimox 500 mg	7	0,7%
lansoprazole hexpharm 30mg	10	1,1%
mefenamic acid sejahtera 500mg	14	1,5%
...	...	...
antasida doen triman	11	1,2%
obh combi batuk plus flu syr rasa menthol 60 ml	9	0,9%
crystalin air mineral 600 ml	12	1,3%
methylprednisolone infion4mg	7	0,7%
mefenamic acid triman 500mg	8	0,8%

Therefore, this value is to ensure that the analysis still covers relevant products, even with a relatively low frequency of transactions. Values that do not meet will then be removed, by removing values that do not meet the minimum support, further analysis will focus on more significant purchase patterns. Products that meet the minimum support value can be seen in the following table.

TABLE 5.  
DATA THAT MEETS THE MINIMUM SUPPORT OF 4%

Product Name	Support
hemaviton stamina plus	4,7%
vicee 500	8,2%
sanmol 500 mg	5,6%
amoxicillin hexpharm 500mg	7,0%
paracetamol trifa 500mg	8,2%

2) 2-Itemset Formation

The 2-itemset frequency pattern is formed by combining all items that meet the minimum criteria from the previous table, then calculating the Support value for each 2-itemset combination produced. The following is an example of 2-itemset Support value calculation:

$$support(A, B) = \frac{Jumlah\ Transaksi\ A,B}{Total\ Transaksi} \times 100\% \dots(2)$$

$$support = \frac{Jumlah\ (Hemaviton, Vicee)}{Total\ Transaksi} \times 100\%$$

$$support = \frac{39}{907} \times 100\%$$

$$support = 4,2\%$$

The results of the Support 2-itemset calculation can be seen in the table 6.

TABLE 6.  
2-ITEMSET SUPPORT RESULT DATA

Product Name	Total Transaction	Support
hemaviton stamina plus, vicee 500	39	4,2%
hemaviton stamina plus, sanmol 500 mg	2	0,2%
hemaviton stamina plus, amoxicillin hexpharm 500mg	0	0%
hemaviton stamina plus, paracetamol trifa 500mg	0	0%
vicee 500, sanmol 500 mg	1	0,1%
vicee 500, amoxicillin hexpharm 500mg	0	0%
vicee 500, paracetamol trifa 500mg	0	0%
sanmol 500 mg, amoxicillin hexpharm 500mg	2	0,2%
sanmol 500 mg, paracetamol trifa 500mg	2	0,2%
amoxicillin hexpharm 500mg, paracetamol trifa 500mg	41	4,5%

Values that do not meet the minimum support will be removed and those that do will be calculated at a later stage.

B. Formation of Association Rules

If all the high-frequency patterns have been found, the next step is to find the association rule that meets the minimum requirement for confidence by calculating the confidence of the associative rule  $A \rightarrow B$ . The association rule is the final result to be achieved, the goal is to select the most relevant rules that can be used as a guide in making decisions and developing more effective strategies. The minimum confidence value is set at 50%, this value is strong enough and reliable for decision making. The following is an example of finding the confidence value of a combination of 2 items:

$$Confidence\ P(A/B) = \frac{Jumlah\ Transaksi\ mengandung\ A,B}{Jumlah\ Transaksi\ A} \times 100\% \dots (3)$$

$$Confidence\ (p) = \frac{Jumlah\ Transaksi\ (Hemaviton, Vicee)}{Jumlah\ Transaksi\ Hemaviton} \times 100\%$$

$$Confidence\ P\left(\frac{A}{B}\right) = \frac{39}{43} \times 100\%$$

$$Confidence\ (P) = 90\%$$

TABLE 7.  
CONFIDENCE VALUE CALCULATION RESULT

Association Rules	Support	Confidence
hemaviton stamina plus, vicee 500	4,2%	90%
amoxicillin hexpharm 500mg, paracetamol trifa 500mg	4,5%	65%

After obtaining the association rules, an evaluation will be carried out using the lift ratio to assess the extent of the

relevance of the rules generated. The following is an example of how to calculate the lift ratio value of the association rule that has been generated:

$$Lift\ Ratio = \frac{Confidence(A,B)}{Support(B)} \times 100\% \dots (4)$$

$$Lift\ Ratio = \frac{Confidence(Hemaviton, Vicee)}{Support(Vicee)} \times 100\%$$

$$Lift\ Ratio = \frac{90}{8,2} \times 100\%$$

$$Lift\ Ratio = 10,9\%$$

The following is a tabulation of the results of the calculation of the lift ratio of the four rules generated :

TABLE 8.  
LIFT RATIO CALCULATION RESULT

Association Rules	Confidence	Lift Ratio
hemaviton stamina plus ==> vicee 500	91%	10,9
amoxicillin hexpharm 500mg ==> paracetamol trifa 500mg	64%	7,8
paracetamol trifa 500mg ==> amoxicillin hexpharm 500mg	53%	7,5
vicee 500 ==> hemaviton stamina plus	52%	11

The lift ratio helps measure how strong the relationship is between two items in an association rule. This lift ratio value provides an indication of the validity of information related to the purchase of item A together with item B. Based on the table above, it is known that the 3 association rules obtained a lift ratio value > 1, then the association rule is declared valid.

C. Testing with Rapidminer Application

In the testing phase with the Rapidminer application, the data used is the same as in the previous manual calculation, but the difference is only in the process, where the data that has gone through the preprocessing stage is directly processed in the Rapidminer application.

1) *The first step* is to drag and drop the imported file into the process area, so that the database operator appears in the main process. After that, add the operators to be used by dragging and dropping them into the process area. Some of the operators used include: 'Numerical to Binominal' and 'W-Apriori'. Next, connect the database with the 'Numerical to Binominal' operator, and then also connect the 'W-Apriori' operator until you reach the final result. Thus, the display of the arrangement of operators used will look like in the following image.

2) *The second stage* is to enter the parameter values that have been determined previously, namely the minimum support value of 4% and the minimum confidence value of 50%.

3) *The next step* is to run the application by selecting the run button.

Based on the test results in Figure 8, it can be seen that the Support 4% and confidence 50% parameters produce 4 rules. The first rule is hemaviton stamina plus ==> vicee 500, the second rule 'amoxicillin hexpharm 500mg ==> paracetamol trifa 500mg, the third rule 'paracetamol trifa 500mg ==> amoxicillin hexpharm 500mg and the fourth rule 'vicee 500 ==> hemaviton stamina plus'.

D. Association Pattern

Based on the tests that have been carried out as a whole, purchases can be analyzed against transaction data at the beleven farma pharmacy which produces the following pattern:

- If you buy hemaviton stamina plus products, you are likely to buy vicee 500 products with a probability value of 91%.
- If you buy amoxicillin hexpharm 500mg, you are likely to buy paracetamol trifa 500mg with a probability value of 64%.
- If you buy paracetamol trifa 500mg product, you are likely to buy amoxicillin hexpharm 500mg product with a probability value of 53%.
- If you buy Vicee 500, you are more likely to buy hemaviton stamina plus with a probability value of 52%.

E. Implication Pattern Results

The results of the implications of this pattern are in the form of a proposed strategy as a result of the analysis, this strategy is prepared to assist the pharmacy in overcoming the problems it faces. By analyzing the existing challenges, this strategy aims to provide solutions so that pharmacies can increase sales through more effective product structuring and make the right product offerings to customers.

1) *An example* of the application of strategies that can be done based on the association pattern generated from the research above, "if customers buy hemaviton stamina plus products, it is recommended to offer vicee products". This is done in order to increase the chances of selling the product.

2) *The resulting* association pattern can be used as information for determining the layout of product sales as one of the strategic efforts made in increasing product sales. Hemaviton products should be placed next to vicee products or it can also be concluded that the placement of health supplement products should be next to vitamin products. The vicee product itself was previously placed on the display shelf at the back of the medicine, the product should be placed next to the hemaviton product which is already on the front of the display shelf. Then the amoxicillin drug is placed adjacent to the paracetamol drug. This is done as an effort to increase customer preferences to make it easier for customers to find what medicine is needed but still in accordance with applicable regulations that the purchase of antibiotic drugs must be based on a doctor's prescription because this drug is not sold over the counter.

3) *From the analysis* conducted, it is also known that the 'curcuma forte' product has a low Support value, meaning

