ASSOCIATION RULE ANALYSIS OF FP-GROWTH ALGORITHM ON DRUG PURCHASE PATTERNS

[1]Dewi Agushinta R., [2]Mega Maralisa Putri
[1,2] Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat
[1]dewiar@staff.gunadarma.ac.id, [2]megamaralisap@gmail.com

Abstrak

Salah satu teknik data mining adalah Association Rule yang sering disebut dengan prosedur Market Basket Analysis untuk mencari pengetahuan tentang pola pembelian konsumen. Riset ini menggunakan algoritma Frequent Pattern Growth (FP-Growth). FP-Growth dalam membentuk itemset dilakukan dengan membuat struktur data FP-Tree. Data yang digunakan memanfaatkan transaksi Klinik selama dua tahun. Analisis dilakukan untuk mempertimbangkan keputusan bagi pemangku kepentingan informasi di Klinik. Hasilnya diperoleh 118 rule dengan nilai minimal 30% support dan 75% confidence. Aturan yang dihasilkan 100% jika beli INJ Piralen maka beli INJ Ranitidine, 100% jika beli Genoint maka beli Genoint SK, jika beli Gastritis Cap maka beli 100% Myalgia Cap, jika beli Sanbe SP NACL Liquid maka beli 100% Glucose Dextrose Liquid, kalau beli Omecidal beli Omedeson 100%, dan kalau beli Omeprazole beli Gludepatik 87%. Riset ini membantu pemilik Klinik menentukan pola dan menampilkan obat yang paling laris dibeli konsumen.

***Kata Kunci**: Aturan Asosiasi, Algoritma FP-Growth, Data Mining*

Abstract

One data mining technique is the Association Rule which is often referred to as Market Basket Analysis procedure to seek knowledge about customer purchasing patterns. This research used Frequent Pattern Growth (FP-Growth) algorithm. FP-Growth in forming itemset is done by creating an FP-Tree data structure. The data utilized the Clinic transaction for two years. The analysis was carried out to consider decisions for information stakeholders at the Clinic. The results obtained 118 rules with a minimum value of 30% support and 75% confidence. The resulting rules are 100% if you buy INJ Piralen then buy INJ Ranitidine, 100% if you buy Genoint then buy Genoint SK, if you buy Gastritis Cap then you buy 100% Myalgia Cap, if you buy Sanbe SP NACL Liquid then you buy 100% Glucose Dextrose Liquid, if you buy Omecidal, you buy Omedeson 100%, and if you buy Omeprazole, you buy Gludepatic 87%. This research helped Clinic owners define the pattern and display the most in-demand drugs purchased to customer.

***Keywords**: Association Rules, Data Mining, FP-Growth Algorithm*

INTRODUCTION

Citra Sehat Clinic is a health care provider with a pharmacy and doctor's practice services. The Clinic located in the Tigaraksa area sells various kinds of medicines. This pharmacy can provide services for purchasing drugs, starting from a doctor's prescription or general medicine needed by the community. In this case, the pharmacy has many transaction data that comes in every year, giving rise to drug purchasing patterns by finding the drug data set that appears most often in a transaction. Therefore, management must know which drugs most frequently purchased and arrange the layout to speed up the transaction process. For example, when a customer buys cough medicine, it will coincide with headache medicine. If the pattern of drug purchases is unknown, it will be hard to determine the layout of the drug, the most frequently purchased drug, and the promotions. Therefore, an analysis of association rules will conduct to define drug patterns purchased.

Data mining is a way of extracting information from a large data set (database). There are many techniques in the data mining process, one of them is the Association Rules. Association Rules are rules for finding high-frequency patterns between sets of items by determining support and confidence. The purpose of data mining is to find association rules from a large database.

Association Rule is one method that aims to find patterns that often appear among many transactions, each transaction consisting of several items [1]. Association Rule is also used in analyzing the Apriori algorithm. A few researchers focused on doing the Association Rule method. Association rules are widely used for the process of analysis and selection. Association Rules have been used in analyzing clay purchase selection [2]. Some are used in the fields of education, sales, transportation, health, and others [3] - [9]. These researches are aimed at seeing interesting knowledge from a large database.

The FP-Growth algorithm which is one of an Association Rules technique will be used. The FP-Growth algorithm is used with a minimum of 30% support and 75% confidence. The purpose of the research is to test the FP-Growth with the Weka application so that the best association rules are obtained. The FP-Growth algorithm, which is the Association Rules technique, can determine the best pattern to use as a reference in obtaining buying patterns. FP-Growth is an algorithm with a strong data structure, resulting in the best opportunity to generate rules. The data used is the transaction from the Clinic at Tigaraksa, Tangerang Banten for two years. This research intends to give the best recommendations for the Clinic.

A. Knowledge Discovery in Database (KDD) Stage

Data mining is one of a series of processes of finding knowledge in a database [10]. Data mining is not a completely new field. One of the difficulties in defining data mining is the fact that data mining inherits many aspects and techniques from established fields of science. The terms data mining and Knowledge Discovery in Database (KDD) are often used interchangeably to describe the process of extracting hidden information in a large database. The KDD process as shown in Figure 1.

Understanding the problems to be faced and having the context to propose viable and tangible solutions is needed. The study of nature, limits, and rules of the data or information and determining the goals to be achieved are necessary. Available data is selected to learn from data sets with predetermined goals. Then integrate them into one data that can help analyze and achieve the target. This information can be found on the same resource or also distributed. In the preprocessing stage, the reliability of the information, namely the implementation of tasks that guarantee the usability of the data, was determined. For this reason, data cleaning, the treatment of missing data, or removing outliers proceeded. It implies omitting variables or attributes with missing data or omitting useless information like text, images, and more. At the data transformation, data quality is transforming into the number reduction of variables in the data set or changing numeric values to categorical ones. In data mining phase, an appropriate data mining process can select whether classification, regression, or clustering, according to the goals set. Furthermore, the two techniques or algorithms are selected to look for patterns and gain knowledge. Meta-learning focuses on explaining why an algorithm works better for a particular problem. There are various possible ways of picking it for each technique. Each algorithm has its essence and its way of working and getting results. Knowing what candidate properties to use and seeing which best fits the data is advisable. Finally, once the technique is selected, the next step is to apply it to the data that has been selected, cleaned, and processed. The execution of the algorithm in some trying parameters may optimize the result. These parameters vary according to the method chosen. Once the algorithms were applied to the data set, the generated patterns were evaluated and the obtained performance was verified. It will meet the goals set in the first phases. There is an evaluation technique called Cross-Validation. It performs data partition, dividing it into training and test data. If all the steps are followed correctly and the evaluation results are satisfied, the last stage is simply applying the knowledge found to the context and beginning to solve its problems. On the contrary, it is necessary to return to the previous stages to make some adjustments, analyzing from the data selection to the evaluation stage.

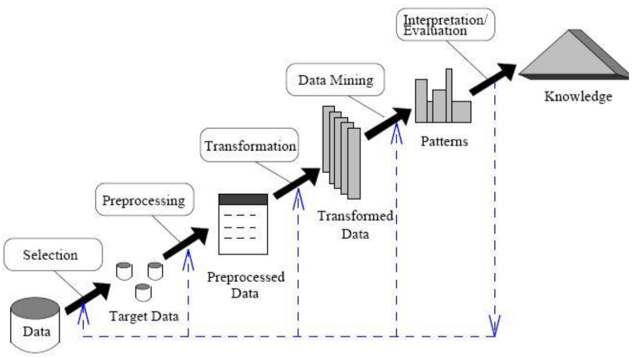


Figure 1. Knowledge Discovery in Database (KDD)

B. Data Mining

Data Mining is the discovery of new information by looking for certain patterns or rules from a large database [1]. Data mining is one of a task series that can be performed, namely [11] :

1. Description

   The process of finding data to describe patterns and trends contained in the data.

2. Estimate

   Estimates are almost the same as classifications, except that the target variable is more numerical rather than categorical. The model is built using a complete record that provides the value of the target variable as a prediction.

3. Prediction

   Prediction is almost the same as classification and estimation, except that in predication the value of the results will be in the future.

4. Classification

   In classification, there are target variable categories.

5. Clustering

   Clustering is a grouping of records, observations, or pay attention and form classes of objects that have similarities.

6. Association

   The task of association in data mining is to find attributes that appear at one time. This stage looks for combinations of items that meet the minimum requirements of value.

C. Association Rule

The association rule will use training data as the notion of data mining, to produce knowledge. Knowledge to find out shopping items that are often bought simultaneously at one time. Association rules in the form of "if ... then ..." is knowledge generated from the association rules function [12].

This stage looked for a combination of items that meet the minimum requirements of the support value in the database. The value of an item's support is obtained by the following formula [2] :

$$\text{Support (A)} = (JT(A))/T \times 100\% \qquad (1)$$

Based on Equation (1), then A is the name of the item, JT (A) the appearance of item A on the transaction, and T is the total transaction. In general, the association rule is a two-step process [13] :

1. Find all frequent itemset, with the understanding that each itemset will often appear with a specified minimum support (min-sup) value.
2. Generate strong association rules from frequent itemset. Understanding these rules must meet minimum support and minimum confidence.

After all, high-frequency patterns have been found, then the association rules that are eligible for confidence are searched by calculating the confidence rules of the association A → B. The confidence value of the A → B rule is obtained from the following formula [2] :

$$\textit{Confidence}\ \mathrm{P}\ (\mathrm{A|B}) = \Sigma \frac{\mathrm{T}ransactions\ Containing\ A\ and\ B}{\mathrm{Transac}tions\ Containing\ \mathrm{A}} \ \mathrm{x}\ 100\% \qquad (2)$$

Based on Equation (2), P(A|B) is an item name, the transaction containing A and B is the appearance of items A and B on the transactions, and the transactions containing A is the appearance of item A on the transaction.

D. FP-Growth Algorithm

FP-Growth is one of the algorithms included in the Rule Mining Association. FP-Growth is an algorithm that has a strong data structure to produce a good rule generating opportunity. FP-Growth is used to determine the best pattern that will be used as a reference in getting purchasing items at the Clinic. Unlike in the Apriori algorithm, there is no need for candidate generation with FP trees, and the frequently occurring item sets are discovered by simply crossing the FP tree [4]. Any set of records valid rules and patterns can be created with this FP tree. FP growth was implemented for book recommendations from a library database. The accuracy given is 60.78% [14].

The FP-Growth algorithm is divided into three main steps [15] :

1. The generation phase of a Conditional Pattern Base, a sub-database that contains a prefix path and a suffix pattern. The generation of a conditional pattern base is obtained through FP-trees that have been built before.
2. Conditional FP-tree generation stage. At this stage the support count of each item in each conditional pattern base is added up, then each item that has support count greater than the minimum support count will be raised with a conditional FP-tree.
3. The frequent itemset search stage if the Conditional FP-tree is a single path, then a frequent itemset is obtained by combining items for each FP-tree conditional.

RESEARCH METHOD

The research method is the sequence of frameworks that must be followed, the order of the flow of this research is a description of the steps that must be passed so that this research can run well. The research method process is as shown in Figure 2.

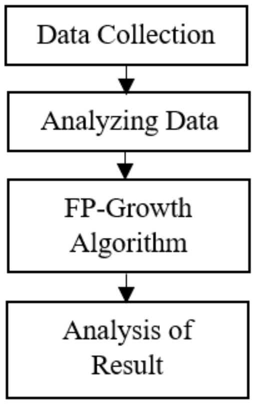


Figure 2. Research Flow

Based on Figure 2, the sequence of research flow is explained as follows :

1. Data Collection used to collect data and information needed in analysis activities. The data collection used was transaction data of a Clinic for two years, 2017-2018.
2. Analyzing Data uses the FP-Growth algorithm. Data processing in the algorithm is as material for analyzing transaction data.
3. Implementation with Weka, a test for all data that will be tested with the Weka application using the FP-Growth algorithm.
4. Results Analysis, the results obtained by the FP-Growth algorithm. At this stage, an analysis of the results obtained from the association rules is formed.

RESULT AND DISCUSSION

A. Data Collection

The data used is transaction data at a Clinic in Tangerang, Banten for two years. Data representation is done by forming transaction patterns every day. If there is the same drug name found then it will be included in the next transaction. There are as many as 81 different kinds of drugs, 700 transactions, with 9407 data on all sales. The drug attributes analyzed in the test were 16 different types of drugs, included in one itemset in manual calculations. Testing will be carried out only on a data sample of 36 transactions with all data of 700 transactions. The data transaction attributes will be the date, number of transactions, the name and the code of the drugs.

B. Analyzing Data

The sample of data to be tested in this study were 81 drugs. There are 36 drug sales transaction data at the Clinic. The name of the drug is initials made to reduce the amount of drug writing so that the drug data is not exchanged. The example drug initials shown are 16 items. Drug initials can be seen in Table 1.

These initial drug names will be used as a sample of 36 transaction data collection. Sample transaction data can be seen in Table 2.

Table 1. Transaction Data

No	Date	Transaction	Drug Name	Drug Code
1	1-01-17	1	L Gula On-Call Plus	LG
2		1	Lanset 100	L100
3		1	INJ Ranitidine	INR
4		1	INJ Piralen	INP
5		1	DR KIR	KIR
6		1	Genoint	GEN
7		1	Genoint SK	GSK
8		2	DR KIR	KIR
9		2	Genoint	GEN
10		2	Genoint SK	GSK
…		…	…	…
9407		700	Bioplacenton	BIO

Table 2. Drug Initials

No	Drugs Name	Initial
1	L Gula On-Call Plus	A
2	Omedical Tab	D
3	Omedeson Tab	E
4	Gludepatic Tab	G
5	Cap Myalgia	J
6	Mikrogynon	T
7	Omeprazole Tab	W
8	Cap Gastristis	Z
9	Cairan NACL Sanbe SP	L_2
10	Genoint	O_2
11	Genoint SK	U_2
12	Glucose Dextrose Liquid	Y_2
13	INJ Ranitidine	A_3
14	INJ Piralen	G_3
15	Salep 24	Q_3
16	Abbocath 24	Z_3

Table 3. Binary Symmetric Tabulation

No	A	B	C	D	E	F	G	…	C_4
1	Y	Y	Y	N	N	N	N	…	N
2	N	N	N	Y	Y	N	N	…	N
3	Y	N	N	Y	Y	N	N	…	N
4	N	N	N	Y	Y	N	Y	…	N
5	Y	N	N	N	N	N	N	…	N
6	N	N	N	Y	Y	N	N	…	N
7	Y	Y	N	Y	Y	N	Y	…	N
8	Y	N	N	Y	Y	Y	Y	…	Y
9	Y	Y	N	N	N	Y	N	…	N
10	N	N	N	Y	Y	Y	Y	…	N
11	Y	N	N	N	N	N	N	…	Y
12	N	N	Y	N	N	N	Y	…	N
13	N	Y	N	N	N	N	N	…	Y
14	N	N	N	N	Y	N	Y	…	N
15	N	N	N	N	N	N	Y	…	Y
16	Y	N	N	Y	Y	N	N	…	N

No	A	B	C	D	E	F	G	…	C_4
17	N	N	N	Y	Y	N	Y	…	N
18	Y	N	N	N	N	N	Y	…	N
19	N	N	N	N	N	N	N	…	N
20	N	N	N	N	N	N	Y	…	N
21	N	Y	Y	N	Y	N	Y	…	N
22	N	N	N	N	Y	N	N	…	N
…	…	…	…	…	…	…	…	…	…
36	N	N	N	N	N	N	N	…	Y

The next step is to form a binary symmetric tabulation to see the appearance of the drug in the transaction. Sample transaction data can be seen in Table 3.

C. FP-Growth Algorithm

The data analysis phase will be carried out using the FP-Growth method. Two inputs are needed to find the rules of the association, determining the minimum support (emergence value) and minimum confidence (trust value). After that, an itemset filter is performed to form an FP-tree, then the association rule process is carried out. Association rule formed must meet the minimum value specified by the user [12]. The author takes a minimum support value of 30% and minimum confidence of 75% from the interview.

These values are the limit of the support and confidence value for the data mining calculation process. This stage looked for 1 itemset that meets the 30% support using the equation (1). After searching for 1 itemset as many as 16 items were obtained, the items are sorted from the highest frequency. Table 4 below is the result of forming 1 itemset.

The obtained items with above 30% support are A, D, E, G, J, T, W, Z, L_2, O_2, U_2, Y_2, A_3, G_3, Q_3, and Z_3. These 16 items will be included in the FP-Tree development transaction. Table 5 follows the highest frequency adjusted transactions.

Table 4. The occurrence of High Frequency Based Items

Initial	Count	Support (%)
Z	23	63,89
J	22	61,11
Y_2	18	50
G_3	18	50
E	17	47,22
W	17	47,22
T	16	44,44
U_2	16	44,44
A_3	16	44,44
G	15	41,67
O_2	15	41,67
D	14	38,89
A	13	36,11
Q_3	13	36,11
Z_3	12	33,33
U_2	16	44,44

Table 5. High-Frequency Adjustable Transaction

TID	Initial	TID	Initial
1	G_3, U_2, A_3, O_2, A	19	Z, J, Y_2, G_3, L_2, A_3
2	Z, J, Y_2, G_3, E, L_2, T, U_2, A_3, O_2, D, Q_3, Z_3	20	Z, J, W, U_2, G, O_2, Q_3, Z_3
3	Z, J, Y_2, G_3, E, W, L_2, A_3, D, A, Q_3, Z_3	21	E, W, T, G
4	Z, J, E, W, T, G, D, Q_3, Z_3	22	Y_2, E, L_2, U_2, O_2, Q_3
5	Z, J, Y_2, G_3, L_2, A_3, A, Q_3	23	G_3, T, U_2, A_3, O_2, Q_3, Z_3
6	Z, J, Y_2, G_3, E, W, L_2, A_3, D	24	G_3, E, T, A_3, G, D
7	Z, J, E, W, T, U_2, G, D, Q_3, Z_3	25	Z, J, E, W, T, G, D
8	E, W, T, U_2, G, O_2, D, A, Q_3, Z_3	26	J, Y_2, G_3, L_2, A_3, A
9	Z, J, Y_2, G_3, L_2, T, U_2, A_3, O_2, A	27	Z, J, G_3, E, A_3, D, A
10	Z, J, Y_2, E, W L_2, G, D	28	Z, J, Y_2, W, L_2
11	Z, J, Y_2, W, L_2, U_2, A, Q_3, Z_3	29	G_3, U_2, A_3, O_2, Q_3, Z_3
12	W, T, U_2, G, O_2, Z_3	30	Z, J, Y_2, G_3, L_2, A_3, A
13	Z, Y_2, G_3, L_2, T, U_2, A_3, O_2	31	Z, J, T, O_2, Q_3
14	Z, Y_2, G_3, E, W, L_2, T, A_3, G	32	Z, J, Y_2, E, D
15	Z, J, G_3, W, U_2, A_3, G, O_2	33	Z, J, Y_2, W, L_2, T, G
16	Z, E, U_2, O_2, D, A, Q_3, Z_3	34	Z, J, G_3, U_2, Z_3
17	Z, J, E, W, T, G, D	35	E, W, T, G, D, A
18	Y_2, L_2, U_2, G, O_2, A	36	Y_2, G_3, L_2, O_2

The establishment of FP-Tree is a compressed data storage structure. FP-Tree is built by mapping each transaction data to each path. FP-Tree is formed by a root that is labeled null. Each node of FP-Tree contains three important information, which is the item label that informs the type of item being presented, the support count represents the number of paths through that node, and the connecting pointers that connect the nodes with the same item label between paths are marked with arrow lines. Thicker knots show the reading of each transaction. The steps to build FP-Tree are as follows :

1. TID Readings 1

   In TID 1 there are items {G_3, U_2, A_3, O_2, A} which then form all path G_3->U_2->A_3->O_2->A. See Figure 3 for more details.

2. TID Readings 2

   After reading TID 1, the next reading TID 2 is {Z, J, Y_2, G_3, E, L_2, T, U_2, A_3, O_2, D, Q_3, Z_3}. In TID 2 you must make a new track again because Z item has never been missed. For more details can be seen in Figure 4.

3. TID Readings 36

   After reading all the TID 1 to TID 2 transactions, the tree form is obtained at the last transaction, which is the 36 FP-Tree TID form. Figure 5 follows is the FP-Tree of all transactions.

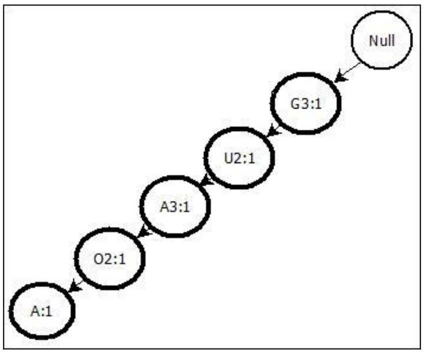


Figure 3. FP-Tree TID 1

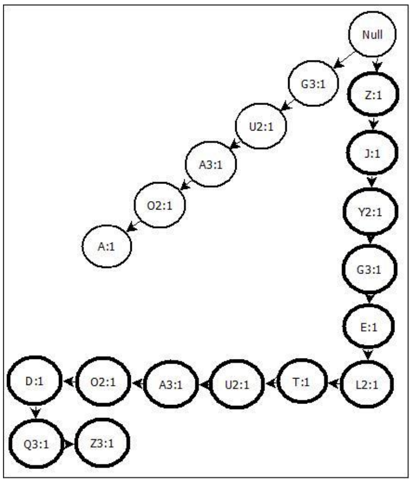


Figure 4. FP-Tree TID 2

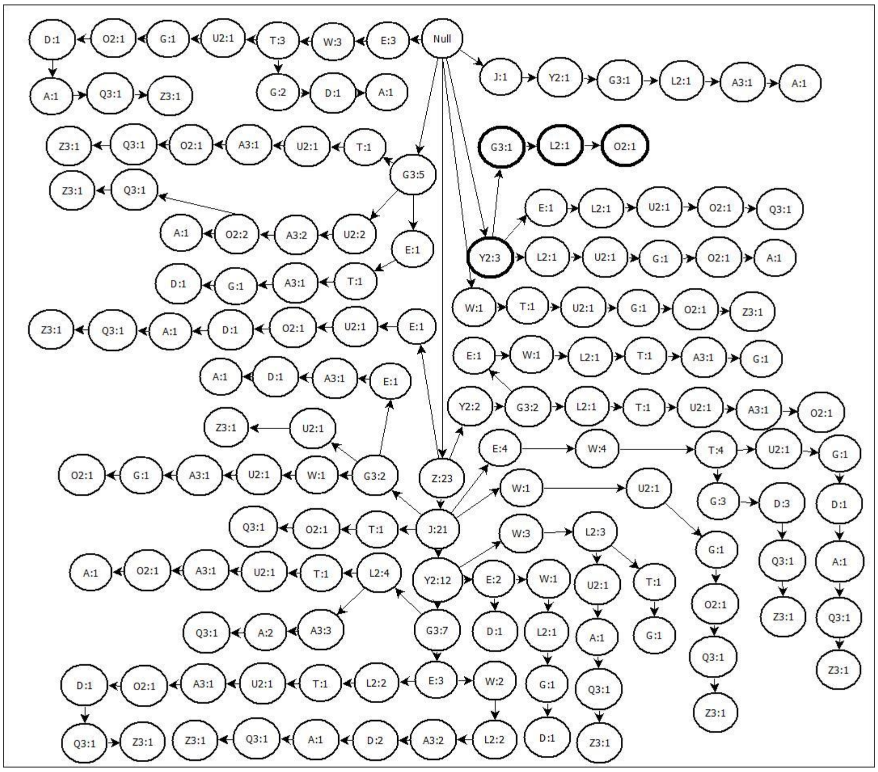


Figure 5. FP-Tree TID 36

In finding frequent itemset, it is necessary to open 36 trees that have been made which are called subtree or tree (trees whose roots are descended from the root of the parent tree). The subtree search starts from the item that has the smallest support (Z_3) to the largest (Z). The process of forming a subtree explains in the following:

- Z_3 Item Track

  All tracks that have Z_3 items are displayed, tracks that do not have Z_3 items are deleted. The reference path taken is the complete FP-Tree on TID 36. The path of Z_3 item can be seen in Figure 6.

- Z Item Track

  The opening process of the FP-Tree tree on Z3 to Z items is the lowest to highest frequency items. Then next to the item with the highest frequency is Z item. All tracks that have Z item is displayed; tracks that do not have Z item are deleted. The reference path taken is the complete FP-Tree on TID 36. The path of Z item can be seen in Figure 7.

Z path has a single node, Z, which has the highest count value. From the FP-Tree generation that has been carried out, the results of Conditional Pattern Base and Conditional FP-Tree can be seen in Table 6.

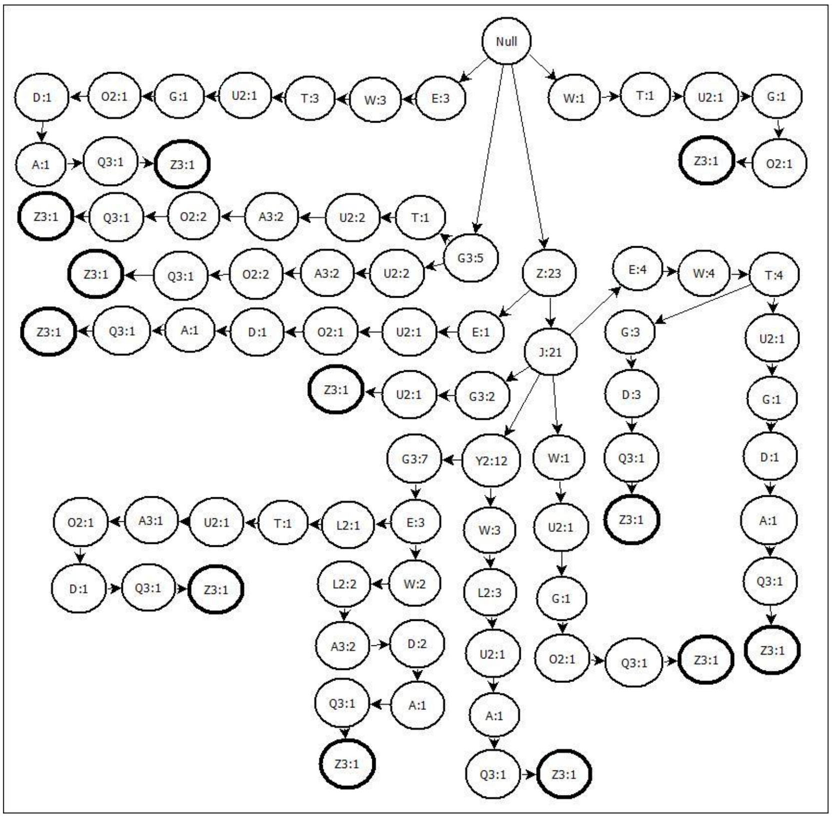


Figure 6. Tracks That Have Z_3 Item

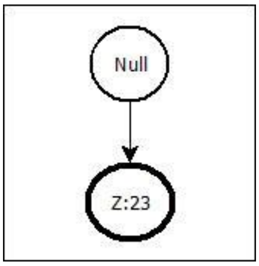


Figure 7. Tracks That Have Z Item

Table 6. Conditional Pattern Base and Conditional FP-Tree

Suffix	Conditional Pattern Base	Conditional FP-Tree
Z_3	{(E, W, T, U_2, G, O_2, D, A, Q_3):1, (G_3, U_2, A_3, O_2, Q_3):1, (G_3, T, U_2, A_3, O_2, Q_3):1, (Z, E, U_2, O_2, D, A, Q_3):1, (Z, J, G_3, U_2):1, (Z, J, Y_2, G_3, E, L_2, T, U_2, A_3, O_2, D, Q_3):1, (Z, J, Y_2, G_3, E, W, L_2, A_3, D, A, Q_3):1, (Z, J, Y_2, W, L_2, U_2, A, Q_3):1, (Z, J, W, U_2, G, O_2, Q_3):1, (Z, J, E, W, T, G, D, Q_3):1, (Z, J, E, W, T, U2, G, D, A, Q_3):1, (W, T, U_2, G, O_2):1}	{(Z:8, J:7, Y_2:3, G_3:5, E:6, W:7, L_2:3, T:6, U_2:10, A_3:4, G:5, O_2:7, D:5, A:5, Q_3:10)}
Q_3	{(E, W, T, U_2, G, O_2, D, A):1, (G_3, U_2, A_3, O_2):1, (G_3, T, U_2, A_3, O_2):1, (Z, E, U_2, O_2, D, A,):1, (Z, J, T, O_2):1, (Z, J, Y_2, G_3, L_2, A_3, A):1 (Z, J, Y_2, G_3, E, L_2, T, U_2, A_3, O_2, D):1, (Z, J, Y_2, G_3, E, W, L_2, A_3, D, A):1, (Z, J, Y2, W, L_2, U_2, A):1, (Z, J, W, U_2, G, O_2):1, (Z, J, E, W, T, G, D):1, (Z, J, E, W, T, U_2, G, D, A):1, (Y_2, E, L_2, U_2, O_2):1}	{(Z:9, J:8, Y_2:5, G_3:5, E:7, W:6, L_2:5, T:6, U_2:9, A_3:5, G:4, O_2:8, D:6, A:6)}
A	{(E, W, T, U_2, G, O_2, D):1, (E, W, T, G, D):1, (G_3, U_2, A_3, O_2):1, (Z, E, U_2, O_2, D):1, (Z, J, G_3, E, A_3, D):1, (Z, J, Y_2, G_3, L_2, T, U_2, A_3, O_2):1, (Z, J, Y_2, G_3, L_2, A_3):1, (Z, J, Y_2, G_3, E, W, L_2, A_3, D):1, (Z, J, Y_2, W, L_2, U_2):1, (Z, J, E, W, T, U_2, G, D):1, (Y_2, L_2, U_2, G, O_2):1, (J, Y_2, G_3, L_2, A_3):1}	{(Z:7, J:6, Y_2:6, G_3:6, E:6, W:5, L_2:6, T:4, U_2:7, A_3:6, G:4, O_2:5, D:4)}
D	{(E, W, T, U_2, G, O_2):1, (E, W, T, G):1, (G_3, E, T, A_3, G):1, (Z, E, U_2, O_2):1, (Z, J, G_3, E, A_3):1, (Z, J, Y_2, E, L_2, A_3):1, (Z, J, Y_2, G_3, E, W, L_2, T, U_2, A_3, O_2):1, (Z, J, Y_2, G_3, E, W, L_2, A_3):1, (Z, J, Y_2, E):1, (Z, J, Y_2, E, W, L_2, G):1, (Z, J, E, W, T, G):1, (Z, J, E, W, T, U_2, G):1}	{(Z:9, J:8, Y_2:5, G_3:4, E:12, W:7, L_2:5, T:6, U_2:4, A_3:5, G:6, O_2:3)}
O_2	{(E, W, T, U_2, G):1, (G3, T, U_2, A_3):1, (G_3, U_2, A_3):1, (Z, E, U_2):1,	{(Z:7, J:5, Y_2:6, G_3:7, E:4, W:4, L_2:6,

https://doi.org/10.35760/tr.2022.v27i3.4626

Suffix	Conditional Pattern Base	Conditional FP-Tree
	(Z, J, G_3, W, U_2, A_3, G):1, (Z, J, T):1, (Z, J, Y2, G_3, L_2, T, U_2, A_3):1, (Z, J, Y_2, G_3, E, L_2, T, U_2, A_3):1, (Z, J, W, U_2, G):1, (Z, J, Y_2, G_3, E, L_2, T, U_2, A_3):1, (Y_2, L_2, U_2, G):1, (Y_2, E, L_2, U_2):1, (Y_2, G_3, L_2):1}	T:6, U_2:12, A_3:6, G:5)}
G	{(E, W, T, U_2):1, (E, W, T):1, (G_3, E, T, A_3):1, (Z, J, G_3, W, U_2, A_3):1, (Z, J, Y_2, E, W, L_2):1, (Z, J, Y_2, W, L_2, T):1, (Z, J, W, U_2):1, (Z, J, E, W, T, U_2):1, (Z, J, Y_2, G_3, E, W, L_2, T, A_3):1, (W, T, U_2):1, (Y_2, L_2, U_2):1}	{(Z:6, J:5, Y_2:4, G_3:3, E:6, W:9, L_2:4, T:7, U_2:6, A_3:3)}
A_3	{(G_3, U_2):1, (G_3, T, U_2):1, (G_3, E, T):1, (Z, J, G_3, E):1, (Z, J, G_3, W, U_2):1, (Z, J, Y2, G_3, L_2):1, (Z, J, Y_2, G_3, L_2, T, U_2):1, (Z, J, Y_2, G_3, E, W, L_2, T, U_2):1, (Z, J, Y_2, G_3, E, W, L_2):1, (Z, Y_2, G_3, L_2, T, U_2):1, (Z, Y_2, G_3, E, W, L_2, T):1, (J, Y_2, G_3):1}	{(Z:8, J:7, Y_2:7, G_3:11, E:5, W:4, L_2:6, T:6, U_2:6)}
U_2	{(E, W, T):1, (G_3):1, (T):1, (Z, E):1, (Z, J, G_3):1, (Z, J, G_3, W):1, (Z, J, Y_2, G_3, L2, T):1, (Z, J, Y_2, G_3, E, L2, T):1, (Z, J, Y_2, W, L_2):1, (Z, J, W):1, (Z, J, E, W, T):1, (Z, Y_2, G_3, L_2, T):1, (W, T):1, (Y_2, L_2):1, (Y_2, E, L_2):1}	{(Z:9, J:7, Y_2:6, G_3:6, E:5, W:6, L_2:6, T:7)}
T	{(E, W):1, (G_3):1, (G_3, E):1, (Z, J):1, (Z, J, Y_2, G_3, L_2):1, (Z, J, Y_2, G_3, E, L_2):1, (Z, J, Y_2, W, L_2):1, (Z, J, E, W):1, (Z, Y_2, G_3, L_2):1, (Z, Y_2, G_3, E, W, L_2):1, (W):1}	{(Z:7, J:5, Y_2:5, G_3:6, E:4, W:5, L_2:5)}
L_2	{(Z, J, Y_2, G_3, E, W):1, (W, J, Y_2, G_3):1, (Z, J, Y2, G_3, E):1, (Z, J, Y_2, W):1, (Z, J, Y_2, E, W):1, (Z, Y_2, G_3):1, (Z, Y_2, G_3, E, W):1, (Y_2):1, (Y_2, E):1, (Y2, G_3):1, (J, Y_2, G_3):1}	{(Z:7, J:6, Y_2:11, G_3:7, E:5, W:4)}
W	{(E):1, (Z, J, G_3):1, (Z, J, Y_2, G_3, E):1, (Z, J, Y_2, E):1, (Z, J, Y_2):1, (Z, J):1, (Z, J, E):1, (Z, Y_2, G_3, E):1	{(Z:7, J:6, Y_2:4, G_3:3, E:5)}
E	{(G_3):1, (Z):1, (Z, J, G_3):1, (Z, J, Y_2, G_3):1, (Z, J):1, (Z, Y_2, G_3):1, (Y_2):1	{(Z:5, J:3, Y_2:3, G_3:4)}
G_3	{(Z, J):1, (Z, J, Y_2):1, (Z, Y_2):1, (Y_2):1, (J, Y_2):1 }	{(Z:3, J:3, Y_2:4)}
Y_2	{(Z, J):1, (J):1 }	{(Z:1, J:2)}
J	{(Z):1}	{(Z:1)}
Z	-	-

Table 6 contains a suffix, conditional pattern base, and conditional FP-Tree. A suffix is the name of the item that is included in the frequent list and is sorted by the smallest to largest support values. Conditional pattern base is the reading of FP-Tree nodes after the formation of TID 36 and is seen according to the suffix of the item, starting from the item with the smallest support value, Z3 item, all paths ending in Z3 are displayed, paths that do not have Z3 item deleted. Z has no path because Z is a single path that has no other vertices, and Z is the item with the largest support value. Conditional FP-Tree is the number of items that appear on the conditional pattern base (on each path of an item) and what items are seen in the reading at each node item.

From the results of the conditional pattern base and conditional FP-Tree, we will get a combination of itemset, the largest item appearing on the FP-Tree conditional then combined with suffix (according to the path of the item). The confidence calculation uses equation (2). Table 7 is a combination of items that meet the value of support and confidence.

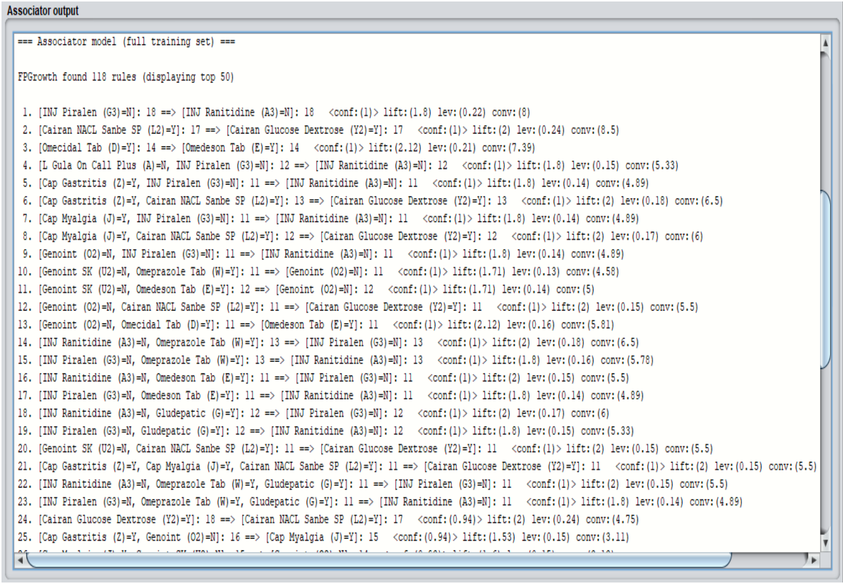
Association rules are obtained by fulfilling minimum support of 30% and 75% confidence. Example of reading the association rules, D => E can be read "if you buy D (Omecidal) then you will buy E (Omedeson) with support 38,89% and 80,85% confidence", and L_2 => Y_2 can be read "if you buy L_2 (NACL liquid Sanbe SP) will then buy Y_2 (Glucose Dextrose Liquid) with 100% confidence".

4. Implementation

Implementation is carried out using the Weka application, with minimum support of 30% and confidence of 75%. The test results obtained 118 rules and it can be seen in Figure 8.

Table 7. Confidence Calculation Results and Association Rules Results

Itemset	Support	Support (%)	Confidence Eq.	Confidence (%)
D => E	14	38,89	{D, E}/D	14/14 = 100
E => D	14	38,89	{E, D}/E	14/17 = 80,85
O_2 => U_2	13	36,11	{O_2, U_2}/O_2	13/15 = 87,80
U_2 => O_2	13	36,11	{U_2, O_2}/U_2	13/16 = 81,81
G => W	13	36,11	{G, W}/G	13/15 = 87,80
W => G	13	36,11	{W, G}/W	13/17 = 81,81
A_3 => G_3	16	44,44	{A_3, G_3}/A_3	16/16 = 100
G_3 => A_3	16	44,44	{G_3, A_3}/G_3	16/18 = 93,61
L_2 => Y_2	17	47,22	{L_2, Y_2}/L_2	17/17 = 100
Y_2 => L_2	17	47,22	{ Y_2, L_2}/Y_2	17/18 = 94
J => Z	20	55,56	{J, Z}/J	20/22 = 90,16
Z => J	20	55,56	{Z, J}/Z	20/23 = 87,30



```
Associator output
=== Associator model (full training set) ===

FPGrowth found 118 rules (displaying top 50)

 1. [INJ Piralen (G3)=N]: 18 ==> [INJ Ranitidine (A3)=N]: 18   <conf:(1)> lift:(1.8) lev:(0.22) conv:(8)
 2. [Cairan NACL Sanbe SP (L2)=Y]: 17 ==> [Cairan Glucose Dextrose (Y2)=Y]: 17   <conf:(1)> lift:(2) lev:(0.24) conv:(8.5)
 3. [Omecidal Tab (D)=Y]: 14 ==> [Omedeson Tab (E)=Y]: 14   <conf:(1)> lift:(2.12) lev:(0.21) conv:(7.39)
 4. [L Gula On Call Plus (A)=N, INJ Piralen (G3)=N]: 12 ==> [INJ Ranitidine (A3)=N]: 12   <conf:(1)> lift:(1.8) lev:(0.15) conv:(5.33)
 5. [Cap Gastritis (Z)=Y, INJ Piralen (G3)=N]: 11 ==> [INJ Ranitidine (A3)=N]: 11   <conf:(1)> lift:(1.8) lev:(0.14) conv:(4.89)
 6. [Cap Gastritis (Z)=Y, Cairan NACL Sanbe SP (L2)=Y]: 13 ==> [Cairan Glucose Dextrose (Y2)=Y]: 13   <conf:(1)> lift:(2) lev:(0.18) conv:(6.5)
 7. [Cap Myalgia (J)=Y, INJ Piralen (G3)=N]: 11 ==> [INJ Ranitidine (A3)=N]: 11   <conf:(1)> lift:(1.8) lev:(0.14) conv:(4.89)
 8. [Cap Myalgia (J)=Y, Cairan NACL Sanbe SP (L2)=Y]: 12 ==> [Cairan Glucose Dextrose (Y2)=Y]: 12   <conf:(1)> lift:(2) lev:(0.17) conv:(6)
 9. [Genoint (O2)=N, INJ Piralen (G3)=N]: 11 ==> [INJ Ranitidine (A3)=N]: 11   <conf:(1)> lift:(1.8) lev:(0.14) conv:(4.89)
10. [Genoint SK (U2)=N, Omeprazole Tab (W)=Y]: 11 ==> [Genoint (O2)=N]: 11   <conf:(1)> lift:(1.71) lev:(0.13) conv:(4.58)
11. [Genoint SK (U2)=N, Omedeson Tab (E)=Y]: 12 ==> [Genoint (O2)=N]: 12   <conf:(1)> lift:(1.71) lev:(0.14) conv:(5)
12. [Genoint (O2)=N, Cairan NACL Sanbe SP (L2)=Y]: 11 ==> [Cairan Glucose Dextrose (Y2)=Y]: 11   <conf:(1)> lift:(2) lev:(0.15) conv:(5.5)
13. [Genoint (O2)=N, Omecidal Tab (D)=Y]: 11 ==> [Omedeson Tab (E)=Y]: 11   <conf:(1)> lift:(2.12) lev:(0.16) conv:(5.81)
14. [INJ Ranitidine (A3)=N, Omeprazole Tab (W)=Y]: 13 ==> [INJ Piralen (G3)=N]: 13   <conf:(1)> lift:(2) lev:(0.18) conv:(6.5)
15. [INJ Piralen (G3)=N, Omeprazole Tab (W)=Y]: 13 ==> [INJ Ranitidine (A3)=N]: 13   <conf:(1)> lift:(1.8) lev:(0.16) conv:(5.78)
16. [INJ Ranitidine (A3)=N, Omedeson Tab (E)=Y]: 11 ==> [INJ Piralen (G3)=N]: 11   <conf:(1)> lift:(2) lev:(0.15) conv:(5.5)
17. [INJ Piralen (G3)=N, Omedeson Tab (E)=Y]: 11 ==> [INJ Ranitidine (A3)=N]: 11   <conf:(1)> lift:(1.8) lev:(0.14) conv:(4.89)
18. [INJ Ranitidine (A3)=N, Gludepatic (G)=Y]: 12 ==> [INJ Piralen (G3)=N]: 12   <conf:(1)> lift:(2) lev:(0.17) conv:(6)
19. [INJ Piralen (G3)=N, Gludepatic (G)=Y]: 12 ==> [INJ Ranitidine (A3)=N]: 12   <conf:(1)> lift:(1.8) lev:(0.15) conv:(5.33)
20. [Genoint SK (U2)=N, Cairan NACL Sanbe SP (L2)=Y]: 11 ==> [Cairan Glucose Dextrose (Y2)=Y]: 11   <conf:(1)> lift:(2) lev:(0.15) conv:(5.5)
21. [Cap Gastritis (Z)=Y, Cap Myalgia (J)=Y, Cairan NACL Sanbe SP (L2)=Y]: 11 ==> [Cairan Glucose Dextrose (Y2)=Y]: 11   <conf:(1)> lift:(2) lev:(0.15) conv:(5.5)
22. [INJ Ranitidine (A3)=N, Omeprazole Tab (W)=Y, Gludepatic (G)=Y]: 11 ==> [INJ Piralen (G3)=N]: 11   <conf:(1)> lift:(2) lev:(0.15) conv:(5.5)
23. [INJ Piralen (G3)=N, Omeprazole Tab (W)=Y, Gludepatic (G)=Y]: 11 ==> [INJ Ranitidine (A3)=N]: 11   <conf:(1)> lift:(1.8) lev:(0.14) conv:(4.89)
24. [Cairan Glucose Dextrose (Y2)=Y]: 18 ==> [Cairan NACL Sanbe SP (L2)=Y]: 17   <conf:(0.94)> lift:(2) lev:(0.24) conv:(4.75)
25. [Cap Gastritis (Z)=Y, Genoint (O2)=N]: 16 ==> [Cap Myalgia (J)=Y]: 15   <conf:(0.94)> lift:(1.53) lev:(0.15) conv:(3.11)
```

Figure 8. Results of the FP-Growth Association Rules

The generated rules from the test are 118. The figure showed only 25 rules, with different levels of confidence. G_3 => A_3 rule can be read as "if you buy G_3 (INJ Piralen) then you will buy A_3 (INJ Ranitidine) with 100% confidence", and L_2 => Y_2 can be read as "if you buy L_2 (Sanbe NACL Liquid) then you will buy Y_2 (Glucose Dextrose Liquid) with 100% confidence".

D. Results Analysis

Data samples are 36 transactions taken from all existing data. From the association rules results, obtained patterns of purchasing drugs are {INJ Piralen, INJ Ranitidine}, {Genoint SK, Genoint}, {Gastritis Cap, Myalgia Cap}, {NACL Sanbe SP Liquid, Dextrose Glucose Liquid}, {Omecidal, Omedeson}, {Omeprazole, Gludepatic}, {Abbocath 24, Ointment 24}, and {Mikrogynon, Gludepatic}. The association rules obtained are the result of the analysis of drug purchase patterns at the Clinic and in Tabel 8.

Table 8. Analysis of Results

If	Then	Confidence (%)
INJ Piralen	INJ Ranitidine	100
Genoint SK	Genoint	100
Gastritis Cap	Myalgia Cap	100
NACL Sanbe SP Liquid	Dextrose Glucose Liquid	100
Omeprazole	Gludepatic	80
Omecidal Tab	Omedeson Tab	100

CONCLUSION AND SUGGESTION

Information was successfully obtained from the results of data mining analysis in the form of drug purchase patterns at the Clinic in Tangerang, Banten with 2 years data collection. The pattern of drug purchases at the Clinic has the highest value of confidence in transactions. The results obtained with a minimum value of 30% support and 75% confidence are 118 rules.

The information obtained can be used by the owner of the Clinic in making decisions in determining the pattern of purchasing the most in-demand drugs. This information can be used to consider the layout decisions of goods, items that are always purchased, and to carry out promotions such as price discounts to customer.

This research is limited to analysis and has not been made in an application. The limitation of the Weka application that could not process large data is another obstacle in completing all existing transaction data. It can also compare with other algorithms such as Apriori and Filtered Associator. It expects to use sales transactions and more attributes to produce more accurate information for further research.

REFERENCES

[1] S. Defit, "Penggunaan Algoritma Apriori Dalam Menganalisa Perilaku Mahasiswa Dalam Memilih Mata Kuliah (Studi Kasus : FKIP UPI YPTK)," *Jurnal Media Processor*, vol. 8, no.3, Oct., pp. 31-42, 2013.

[2] P. Mulya, "Analisa dan Implementasi Association Rule Dengan Algoritma FP-Growth Dalam Seleksi Pembelian Tanah Liat (Studi Kasus Di PT. Anveve Ismi Berjaya)," *Jurnal Teknologi dan Sistem Informasi Bisnis*, vol. 1, no. 1, Jan., pp. 47-57, 2019.

[3] Fadli, M. I. Zulfa, and Y. Ramadhani, "Perbandingan Unjuk Kerja Algoritme Klasifikasi Data Mining dalam Sistem Peringatan Dini Ketepatan Waktu Studi Mahasiswa," *Jurnal Teknologi dan Sistem Komputer*, vol. 6, no. 4, Oct., pp. 158-163, 2018.

[4] S. Sidhu, U. K. Meena, A. Nawani, H. Gupta, and N. Thakur, "FP Growth Algorithm Implementation," *International Journal of Computer Applications*, vol. 93, no. 8, May, pp. 6-10. 2014.

[5] X. Huang, Y. Xu, S. Zhang, and W. Zhang, "Association Rule Mining for Selecting Proper Students to Take Part in Proper Discipline Competition: A Case Study of Zhejiang University of Finance and Economics," *International Journal of Emerging Technologies in Learning*, vol. 13, no. 3, Mar., pp. 100-113, 2018.

[6] N. Sağın and B. Ayvaz, "Determination of Association Rules with Market Basket Analysis: Application in the Retail Sector," *Southeast Europe*

Journal of Soft Computing, vol. 7, no.1, Mar., pp. 10-19, 2018.

[7] R. Riszky and M. Sadikin, "Data Mining Menggunakan Algoritma Apriori Untuk Rekomendasi Produk Bagi Pelanggan", *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 3, Jul., pp. 103-108, 2019.

[8] Ruswati, A. I. Gufroni, and Rianto, "Associative Analysis Data Mining Pattern Against Traffic Accidents Using Apriori Algorithm," *Scientific Journal of Informatics*, vol. 5, no. 2, Nov., pp. 91-104, 2018.

[9] H. Yang, J. H. Kang, Y. B. Park, Y. J. Park, H. S. Oh, and S. B. Kim, "Association Rule Mining and Network Analysis in Oriental Medicine," *PLoS ONE*, vol. 8, no.3, Mar., pp. 1-7, 2013.

[10] O. Maimon and L. Rokach, Data Mining and Knowledge Discovery Handbook, 2nd ed, US: Springer, 2010.

[11] Kusrini and E. T. Luthfi, Algoritma Data Mining, Yogyakarta: Andi, 2009.

[12] K. Tampubolon, H. Saragih, and B. Reza, "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan", *Informasi dan Teknologi Ilmiah* (INTI), Oct., pp. 93-106, 2013.

[13] J. Han, M. Kamber, and J. Pei, Data Mining: Concept and Techniques, 3rd ed, San Fransisco, CA: Morgan Kaufmann, 2012.

[14] R. N. Miraldi, A. Rachmat, and B. Susanto, "lmplementasi Algoritma FP-GROWTH Untuk Sistem Rekomendasi Buku di Perpustakaan UKDW," *INFORMATIKA*, vol. 10, no. l, pp. 29-39, 2014.

[15] S. O. Obaje, J. I. Omade, and U. A. Dambatta, "Clays and Their Industrial Applications: Synoptic Review," *International Journal of Science and Technology*, vol.3, pp. 264-270, 2013