APRIORI ALGORITHM FOR THE DETERMINATION OF THE GOODS SALES MARKET BANK

Fricles Ariwisanto Sianturi[1], Petti Indrayati Sijabat[2], Amran Sitohang[3], R. Mahdalena Simanjorang[4]

Informatics Management Study Program[1]
Technical Informaric Study Program[2]
Technical Informaric Study Program[3]
Informatics Management Study Program[4]
[1,2,3,4]STMIK Pelita Nusantara Medan, North Sumatra, Indonesia
sianturifricles@gmail.com[1], petti.jabat@gmail.com[2], amranryan89@gmail.com[3],
lenasinaga30@gmail.com 4

Abstract

Data mining a process of finding meaningful new relationships, patterns, and trends by filtering the huge data stored in the database using pattern recognition techniques. One of the data mining techniques is the a priori algorithm. A priori algorithm is defined as an algorithm for finding the highest frequency patterns. Currently, the a priori algorithm has been implemented in various fields, one of which is in the field of business or trade and the field of education. Market basket analysis technique or market basket analysis is a data mining technique that aims to find products that are often purchased simultaneously from transaction data.Bina Karya Swalayan is a modern market that has various types of goods. Where in the supermarket there are still some problems faced by a manager and employees.
Keywords: Data Mining, Supermarket, Bina Karya Swalayan

1. Introduction

Data mining has been widely applied in various fields, one of which is in the field of sales, namely the use of the a priori method for market basket analysis in minimarket sales transactions using Java and Mysql. Data mining is the process of extracting knowledge, something important or interesting from the data in the database so that it produces very valuable information. The market basket analysis technique is a technique that adapts data mining science. The goal is to find a relationship (association) between items from a transaction database or relational database based on association rules using the apriori algorithm to assist developers in designing marketing strategies for goods in minimarkets.[1]

Bina Karya Swalayan is a modern market that has various types of goods. Where in the supermarket there are still some problems faced by a manager and employees.

Managers and employees still find it difficult to know what items are often purchased simultaneously by consumers. So it becomes a problem for supermarkets if it cannot be used as information. With the support of technological developments, the ability to collect and process data is also growing. Therefore we need an application that is able to sort and select large data so that useful information can be obtained for its users.

2. Literature Review

2.1 Implementation

In general, the term implementation in the large Indonesian dictionary means implementation. The term implementation is usually associated with an activity carried out to achieve certain goals. Webster's dictionary, in short, states that to implement means to provide the means for carrying out, to give practical effect to (causing an impact or effect on something). This understanding means that to implement something must be accompanied by supporting facilities which will later have an impact or effect on something.[2]

2.2 Data Mining

Data mining or often referred to as knowledge discovery in database (KDD) is an activity that includes the collection, use of historical data to find regularities, patterns, or relationships in large data. This data mining expenditure can be used to help make decisions in the future. The development of KDD causes the use of pattern recognition to decrease because it has become part of data mining[3].

The terms data mining and knowledge discovery in database (KDD) is often used to take turns to explain the excavation process of hidden information in a database big. Both terms have a different concept, but it has interrelation with each other, which is where the stages throughout the Knowledge Discovery in Database (KDD) is a data mining KDD process sizes are as follows:[3]
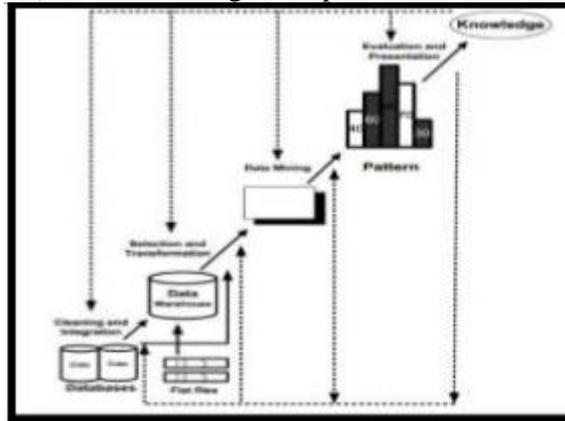

Figure 1. Knowledge Discovery stage on Data Mining (KDD)[4]

Data mining, also known as knowledge discovery in database (KDD), is an activity that involves collecting, using historical data to find regularities, patterns, or relationships in large data sets. The output of data mining can be used to improve future decision making[4].

Important things related to data mining are:

1. *Data mining* is an automatic process of existing data.
2. The data to be processed is in the form of very large data.
3. The purpose of data mining is to find relationships or patterns that may provide useful indications.
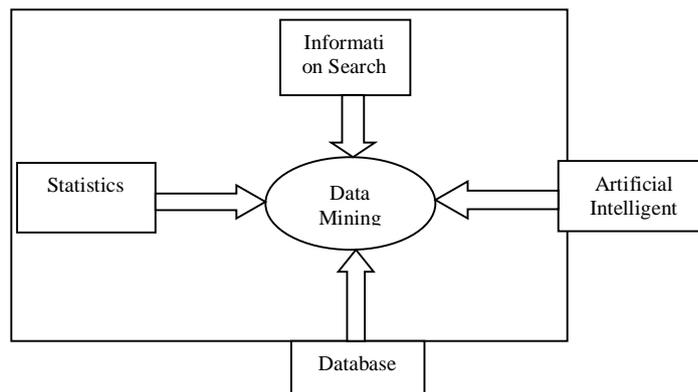

Figure 2. Data Mining Science[4]

2.3 Data Mining Grouping

Data mining divided into groups based on the tasks that can be done[5].

1. Description

Sometimes researchers and analysts simply want to try to find data to describe the patterns and trends contained in the data. For example, polling officials may not be able to determine the statement or the fact that those who are not professional enough will have little support in the presidential election. Descriptions of patterns and trends often provide possible explanations for a pattern or trend.

2. Estimate

   Estimation is almost the same as classification, except that the estimated target variable is more numerical than categorical. The model is built using a complete record that provides the value of the target variable as a prediction. Furthermore, in the next review, the estimated value of the target variable was made based on the value of the prediction variable. For example, we will estimate systolic blood pressure in hospital patients based on patient age, sex, body weight index, and blood sodium level. The relationship between systolic blood pressure and the value of the predictive variable in the learning process will produce an estimation model. The resulting estimation model can be used for other new cases.

3. Prediction.

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

   Examples of business and research predictions are:
   a. Rice price prediction in the next three months.
   b. Predict the percentage increase in traffic accidents next year if the lower speed limit is raised.
   Several methods and techniques used in classification and estimation can also be used (under appropriate circumstances) for prediction.

4. Classification

   In classification, there are target categorical variables. For example, the classification of income can be separated into three categories, namely high income, medium income, and low income.

   Other examples of classifications in business and research are:
   a. Determine whether a credit card transaction is fraudulent or not.
   b. Estimating whether a client's mortgage application is a good or bad credit.
   c. Diagnose a patient's disease to get into what category of disease.

5. Clustering

   Computing is a grouping of records, observations, or attention and forms a class of similar objects. A cluster is a collection of records that are similar to one another and have no resemblance to records in other clusters. Clustering is different from classification in that there is no target variable in the clustering. Clustering does not attempt to classify, estimate, or predict the value of the target variable. However, the clustering algorithm tries to divide the entire data into groups that are similar (homogeneous), where the similarity in one group will be of maximum value, while similarity to records in other groups will be of minimal value.[6]

   Examples of clustering in business and research are:
   a. Get consumer groups for target marketing of a product for companies that do not have large enlargement funds.
   b. For an accounting audit, namely to separate good and suspicious financial behavior.
   c. Doing clustering of the expression of genes, to get similar behavior of genes in large numbers.

6. Association.

   The task of association in data mining is to find attributes that appear at one time. In the business world, it is more commonly called shopping cart analysis.

   Examples of associations in business and research are[1]:
   a. Researching the number of subscribers of cellular telecommunications companies who are expected to give a positive response to the service upgrade offers provided.

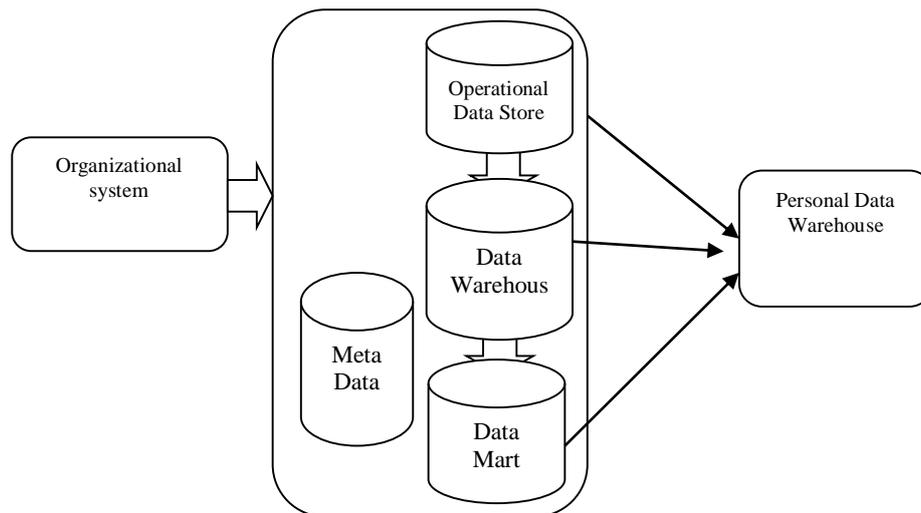b. Determine which supermarket items were purchased simultaneously and which were never purchased simultaneously

Figure 3. Flow of Data in the Formation of a Data Warehouse[7]

2.4 *Association Rule*

Association analysis is also known as a data mining technique which is the basis for various other data mining techniques. In particular, one of the stages of association analysis called high frequency pattern analysis (frequent pattern mining) has attracted the attention of many researchers to produce an efficient algorithm.[8]

The basic methodology of association analysis is divided into two stages:
1. High frequency pattern analysis

This stage looks 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:

2. Establishing associative rules

After all the high frequency patterns are found, then we find an associative rule that meets the minimum requirements for *confidence* by calculating the confidence of the associative rule A » *B* The confidence value of rule A is obtained from the following formula:» *B*

2.5 **Apriori Algorithm**

The priori algorithm is the best known algorithm for finding high frequency patterns. a priori algorithm is divided into several stages called narration or pass.[9][10]
1. The formation of the itemset candidate, the k-itemset candidate is formed from the (k-1) -itemset combination obtained from the previous iteration. One way of the a priori algorithm is the trimming of k-itemset candidates whose subsets containing k-1 items are not included in the high frequency pattern with length k-1.

2. Support calculation for each k-itemset candidate. Support from each k-itemset candidate is obtained by scanning the database to calculate the number of transactions that contain all the items in the k-itemset candidate. This is also a feature of the a priori algorithm where it is necessary to calculate the entire database as many as the longest k-itemset.
3. Set high frequency pattern. High frequency patterns containing k items or k-itemset are determined from k-itemset candidates whose support is greater than the minimum support.
4. If no new high frequency patterns are found then the whole process is terminated. If not, then k is added by one and returns part 1.

3. Results and Discussion

With the case study at Supermarkets Bina Karya Medan, it can be done to analyze the data, especially sales data (cosmetic goods data) with one of the objectives is to find a combination pattern of sales of goods and the relationship between items in the type of goods in the transaction. The data on goods that will be sampled are taken from the goods data in 2013. The following is the item data at Bina Karya Supermarkets.[11]

Table 1. Data on Sales of Goods at Bina Karya Supermarkets Period I of 2013

NO.	Invoice No.	Name of goods	amount
		PERIOD I (JANUARY TO APRIL)	
1	060801	Attack Softener 50gr	96
2	010251	Lifebuoy Bw Btl 300 ml Green	75
3	050513	Molto Ultra 1x Rinse	165
4	140183	Rexona Women Ro Wht 40 ml	26
5	080102	Pepsodent White 75gr	10
6	041854	Fantastic lemon pump 500 ml	9
7	081439	Listerine 500ml Natural Green	15
8	120167	Citra Hbl 120ml Fresh Radianc	14
9	120421	Ponds Am Regen Day Cream 25 ml	16
10	023604	Shinzui Soap 100gr Hana	24
11	124717	Herborist Luur Bali 100 ml Scru	9
12	070101	Dahlia Napht 25gr K-19	11
13	030717	Economy White 500k 300gr	19
14	060143	Rinso Anti Stain 900gr	17
15	080203	Close Up Green 160gr	14
16	160129	Sunslik Shp 10 ml Black Shine	23
17	123020	Jb Lotion 100 ml mlk	26
18	120203	Vaseline Hbl 100 ml Hlty Spf	9
19	160928	Pantene Shp 170 ml AD	20
20	123108	Cussons H Lot 50 ml Natural	7
21	060816	Attact Color 50gr	4
22	102800	My B Powder 50gr Chamomile	-
23	122082	Marina Hbl 100ml Ginseng	8
24	123612	Purbasari LLR 135gr Bengkoan	9
25	060513	Soklin Liq 30 ml Antibacteria	-
26	021800	Kelly lemon soap 15 g	-
27	120856	Nivea multi protecting ff 50 ml	13
28	220259	Pampers orange taped m22	35
29	123615	Purbasari soap 100gr ex betel	17

| 30 | 101203 | Zwitsal Powder 100gr Ext Care | 6 |

Table 2. Data on Sales of Goods at Bina Karya Supermarkets Period II of 2013
PERIOD II (MAY TO AUGUST)

NO.	Invoice No.	Name of goods	amount
1	060801	Attack Softener 50gr	105
2	010251	Lifebuoy Bw Btl 300 ml Green	73
3	050513	Molto Ultra 1x Rinse	98
4	140183	Rexona Women Ro Wht 40 ml	24
5	080102	Pepsodent White 75gr	21
6	041854	Fantastic lemon pump 500 ml	-
7	081439	Listerine 500ml Natural Green	18
8	120167	Citra Hbl 120ml Fresh Radianc	8
9	120421	Ponds Am Regen Day Cream 25 ml	16
10	023604	Shinzui Soap 100gr Hana	24
11	124717	Herborist Luur Bali 100 ml Scru	5
12	070101	Dahlia Napht 25gr K-19	9
13	030717	Economy White 500k 300gr	7
14	060143	Rinso Anti Stain 900gr	11
15	080203	Close Up Green 160gr	14
16	160129	Sunslik Shp 10 ml Black Shine	19
17	123020	Jb Lotion 100 ml mlk	9
18	120203	Vaseline Hbl 100 ml Hlty Spf	12
19	160928	Pantene Shp 170 ml AD	18
20	123108	Cussons H Lot 50 ml Natural	-
21	060816	Attack Color 50gr	9
22	102800	My B Powder 50gr Chamomile	-
23	122082	Marina Hbl 100ml Ginseng	8
24	123612	Purbasari LLR 135gr Bengkoan	8
25	060513	Soklin Liq 30 ml Antibacteria	10
26	021800	Kelly lemon soap 15 g	5
27	120856	Nivea multi protecting ff 50 ml	1
28	220259	Pampers orange taped m22	25
29	123615	Purbasari soap 100gr ex betel	13
30	101203	Zwitsal Powder 100gr Ext Care	-

Table 3. Data on Sales of Goods at Bina Karya Supermarkets Period III 2013
PERIOD III (SEPTEMBER TO DECEMBER)

NO.	Invoice No.	Name of goods	amount
1	060801	Attack Softener 50gr	87
2	010251	Lifebuoy Bw Btl 300 ml Green	35
3	050513	Molto Ultra 1x Rinse	56
4	140183	Rexona Women Ro Wht 40 ml	19
5	080102	Pepsodent White 75gr	33
6	041854	Fantastic lemon pump 500 ml	4
7	081439	Listerine 500ml Natural Green	14
8	120167	Citra Hbl 120ml Fresh Radianc	11
9	120421	Ponds Am Regen Day Cream 25 ml	16

10	023604	Shinzui Soap 100gr Hana	24
11	124717	Herborist Luur Bali 100 ml Scru	5
12	070101	Dahlia Napht 25gr K-19	9
13	030717	Economy White 500k 300gr	7
14	060143	Rinso Anti Stain 900gr	11
15	080203	Close Up Green 160gr	6
16	160129	Sunslik Shp 10 ml Black Shine	14
17	123020	Jb Lotion 100 ml mlk	2
18	120203	Vaseline Hbl 100 ml Hlty Spf	7
19	160928	Pantene Shp 170 ml AD	10
20	123108	Cussons H Lot 50 ml Natural	7
21	060816	Attact Color 50gr	4
22	102800	My B Powder 50gr Chamomile	-
23	122082	Marina Hbl 100ml Ginseng	8
24	123612	Purbasari LLR 135gr Bengkoan	4
25	060513	Soklin Liq 30 ml Antibacteria	-
26	021800	Kelly lemon soap 15 g	5
27	120856	Nivea multi protecting ff 50 ml	2
28	220259	Pampers orange taped m22	20
29	123615	Purbasari soap 100gr ex betel	2
30	101203	Zwitsal Powder 100gr Ext Care	7

4. Analysis And Discussion

Table 4. Candidates for Association Rules from Table 3.

No.	Rules	Support (itemsets)	Support	Confidence (%)
1	If you buy Attack Softener Nt 25gr: you will buy 500ml Natural Green Listerine	12	16	75%
2	If you buy 500ml Natural Green Listerine: then you will buy an Attack Softener Nt 25gr	12	14	85.7%
3	If you buy Attack Softener Nt 25gr: it will buy Molto Ultra 1x Rinse	16	16	100%
4	If you buy Molto Ultra 1x Rinse: you will buy Attack Softener Nt 25gr	16	17	94.1%
5	If you buy Attack Softener Nt 25gr: you will buy Pepsodent White 75gr	15	16	93.7%
6	If you buy Pepsodent White 75gr: you will buy Attack Softener Nt 25gr	15	16	93.7%
7	If purchased, Citra Hbl 120ml Fresh	11	12	91.6%

Table 5. Association Rules That Meet the Minimum Confidence Value from table 4

No.	Rules	Support (%)	Confidence (%)
1	If you buy Attack Softener Nt 25gr: you will buy 500ml Natural Green Listerine	66%	75%
2	If you buy 500ml Natural Green Listerine: then you will buy an Attack Softener Nt 25gr	66%	85.7%
3	If you buy Attack Softener Nt 25gr: it will buy Molto Ultra 1x Rinse	88%	100%

No.	Rules	Support	Confidence
4	If you buy Molto Ultra 1x Rinse: you will buy Attack Softener Nt 25gr	88%	94.1%
5	If you buy Attack Softener Nt 25gr: you will buy Pepsodent White 75gr	83%	93.7%
6	If you buy Pepsodent White 75gr: you will buy Attack Softener Nt 25gr	83%	93.7%
7	If you buy 120ml Fresh Radianc Hbl Citra: it will buy Molto Ultra 1x Rinse	61%	91.6%
8	If you buy Citra Hbl 120ml Fresh Radianc: then you will buy 25ml Ponds Am Regen Day Cream	61%	91.6%
9	If you buy Ponds Am Regen Day Cream 25ml: you will buy Citra Hbl 120ml Fresh Radianc	61%	91.6%
10	If you buy Lifebuoy Bw bottle 300 ml Green: you will buy Molto Ultra 1x Rinse	66%	92.3%

The next formation of association rules is formed from a combination pattern of three itemsets from table 4.11 which meets the minimum support, prospective association rules as in table 4.16 below:

Table 5. Candidates for Association Rules from Table 4

No.	Rules	Support (itemsets)	Support	Confidence (%)
1	If you buy Listerine 500ml Natural Green: then you will buy Molto Ultra 1x Rinse and Attack Softener Nt 25gr	12	14	85.7%
2	If you buy Molto Ultra 1x Rinse: you will buy Attack Softener Nt 25gr and Listerine 500ml Natural Green	12	17	70.5%
3	If you buy Attack Softener Nt 25gr: you will buy Listerine 500ml Natural Green and Molto Ultra 1x Rinse	12	16	75%
4	If you buy Listerine 500ml Natural Green and Molto Ultra 1x Rinse: it will buy Attack Softener Nt 25gr	12	13	92.3%
5	If you buy Listerine 500ml Natural Green and Attack Softener Nt 25gr: you will buy Molto Ultra 1x Rinse	12	12	100%
6	If you buy Molto Ultra 1x Rinse and Attack Softener Nt 25gr: you will buy Listerine 500ml Natural Green	12	16	75%
7	If you buy Listerine 500ml Natural Green: you will buy Pepsodent White 75gr and Attack Softener Nt 25gr	11	14	78.5%

Table 6. Association Rules that Meet the Minimum Confidence Value from Table 5.

No.	Rules	Support (%)	Confidence (%)
1	If you buy Listerine 500ml Natural Green: then you will buy Molto Ultra 1x Rinse and Attack Softener Nt 25gr	66%	85.7%
2	If you buy Molto Ultra 1x Rinse: you will buy Attack Softener Nt 25gr and Listerine 500ml Natural Green	66%	70.5%

3	If you buy Attack Softener Nt 25gr: you will buy Listerine 500ml Natural Green and Molto Ultra 1x Rinse	66%	75%
4	If you buy Listerine 500ml Natural Green and Molto Ultra 1x Rinse: it will buy Attack Softener Nt 25gr	66%	92.3%
5	If you buy Listerine 500ml Natural Green and Attack Softener Nt 25gr: you will buy Molto Ultra 1x Rinse	66%	100%
6	If you buy Molto Ultra 1x Rinse and Attack Softener Nt 25gr: you will buy Listerine 500ml Natural Green	66%	75%
7	If you buy Listerine 500ml Natural Green: you will buy Pepsodent White 75gr and Attack Softener Nt 25gr	61%	78.5%
10	If you buy Pepsodent White 75gr and Attack Softener Nt 25gr: you will buy Listerine 500ml Natural Green	61%	73.3%
11	If you buy Listerine 500ml Natural Green Rinse: you will buy Pepsodent White 75gr and Molto Ultra 1x	61%	78.5%
12	If you buy Listerine 500ml Natural Green and Pepsodent White 75gr: then you will buy Molto Ultra 1x Rinse	61%	91.6%
13	If purchased Listerine 500 ml Natural Green and Molto	61%	84.6%
14	If you buy Pepsodent White 75gr and Molto Ultra 1x Rinse: you will buy 500ml Natural Green Listerine	61%	73.3%
15	If you buy Molto Ultra 1x Rinse: you will buy Pepsodent White 75gr and Attack Softener Nt 25gr,	83%	88.2%
16	If you buy Pepsodent White 75gr: you will buy Attack Softener Nt 25gr and Molto Ultra 1x Rinse	83%	93.7%
17	If you buy Attack Softener Nt 25gr: you will buy Molto Ultra 1x Rinse and Pepsodent White 75gr	83%	93.7%
18	If you buy Molto Ultra 1x Rinse and Pepsodent White 75gr: then you will buy Attack Softener Nt 25gr,	83%	88.2%
19	If you buy Molto Ultra 1x Rinse and Attack Softener Nt 25gr: you will buy Pepsodent White 75gr	83%	93.7%
20	If you buy Pepsodent White 75gr and Attack Softener Nt 25gr: you will buy Molto Ultra 1x Rinse	83%	100%

Table 7. Candidate association rules from table 6.

No.	Rules	Support (itemsets)	Support	Confidence (%)
1	If you buy Listerine 500ml Natural Green: you will buy Pepsodent White 75gr, Attack Softener Nt 25gr and Molto Ultra 1x Rinse	11	14	78.5%
2	If you buy Pepsodent White 75gr: it will buy Attack Softener Nt 25gr, Molto Ultra 1x Rinse and Listerine 500ml Natural Green	11	16	68.7%
3	If you buy Attack Softener Nt 25gr: then you will buy Molto Ultra 1x Rinse, Listerine 500ml Natural Green and Pepsodent White 75gr	11	16	68.7%
4	If you buy Molto Ultra 1x Rinse: you will buy Listerine 500ml Natural Green, Pepsodent White 75gr and Attack Softener Nt 25gr	11	17	64.7%

Table 8. Association Rules that Meet the Minimum Confidence Value from table 7

No.	Rules	Support (%)	Confidence (%)
1	If you buy Listerine 500ml Natural Green: you will buy Pepsodent White 75gr, Attack Softener Nt 25gr and Molto Ultra 1x Rinse	61%	78.5%

1. Conclusion

From the experimental results that the author conducted on this research, the writer can draw some conclusions related to the research process and the content of the research itself. The process must be carried out with complete data so that the assessment will be carried out objectively.The application of classification methods can facilitate and provide settlement calculations and criteria in assessing customer loyalty.

Reference

[1]   H. Sulastri and AI Gufroni, "Application of Data Mining in Grouping Thalassemia Patients," *J. Technol. and Sist. Inf.*, vol. 03, no. 02, pp. 299–305, 2017, doi: 10.25077 / TEKNOSI.v3i2.2017.299-305.

[2]   DK Pane, "Implementation of Data Mining in Electronic Product Sales Using the Apriori Algorithm (Case Study: Kreditplus)," *Pelita Inform. Budi Darma*, 2013.

[3]   M. Fahmi and FA Sianturi, "ANALYSIS OF APRIORIC ALGORITHM ON CONSUMER ORDERING AT CAFÉ THE L. COFFE COFFEE, "*SAINTEK (Journal of Science and Technology.*, vol. 1, no. 1, pp. 52–57, 2019.

[4]   MS Fricles Ariwisanto Sianturi, "COMBINATION OF IMPLEADDITIVEWEIGHTING (SAW) METHODS WITH NEAREST NEIGHBOR ALGORITHM FOR EMPLOYEE RECRUITMENT," *Mantik Penusa*, vol. 3, no. 2, pp. 1–9, 2019, doi: .1037 // 0033-2909.I26.1.78.

[5]   FA Sianturi, T. Informatics, and S. Utara, "Application of the Apriori Algorithm for Level Determination," *Mantik Penusa*, vol. 2, no. 1, pp. 50–57, 2018, [Online]. Available: http://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/330.

[6]   H. Toivonen, "Apriori Algorithm," in *Encyclopedia of Machine Learning and Data Mining*, 2017.

[7]   P. Esling and C. Agon, "Time-series data mining," *ACM Comput. Surv.*, 2012, doi: 10.1145 / 2379776.2379788.

[8]   HD Hutahaean, B. Sinaga, and AA Rajagukguk, "Analysis and Design of Apriori Algorithm Application for Product Sales Correlation (Case Study: Diory Farma Pharmacy)," *JIPN (Journal of Informatics Pelita Nusantara)*, 2016.

[9]   M. Eka and RA Siregar, "DATA MINING ASSOCIATION RULE TO SUPPORT TAX TAX ACHIEVEMENT SYSTEMS IN CASE STUDY AREAS IN DELI SERDANG DISTRICT," *KOMIK (Nas. Technol. Inf. And Computer Conference)*, 2018, doi: 10.30865 / Komik.v2i1.979.

[10]   DS Kusumo, MA Bijaksana, and D. Darmantoro, "DATA MINING WITH APRIORIC ALGORITHM IN ORACLE RDBMS," *TECHNOLOGY - J. Researcher. and Pengemb. Telecomun. Control, Computer, Electr. and Electrons.*, 2016, doi: 10.25124 / tektrika.v8i1.215.

[