

Apriori Algorithm Implementation to Determine Product Sales Priority

Anggit Dwi Hartanto*¹, Bobby Candra Lim², Deas Pradana³

^{1,2,3,4,5} Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta

Email : ¹anggit@amikom.ac.id, ²bobbycandra@gmail.com, ³deaspradana98@gmail.com,

Abstract

Nowadays there are so many stores that stand where every day you have a transaction on selling goods. Sometimes transactions per day are not small, these many transactions can lead to a buildup of data. However, it is rare for business owners to realize that the data can be processed into various kinds of information which can later be used for the interests of companies and stores. The solution to this can be to use an algorithm that is often used today, namely the Apriori algorithm. The Apriori algorithm is one of the data mining techniques used to find high associative patterns between products. The calculation is done by determining the support and confidence between the itemset set in the transaction or also called the association rule function. The test results with a priori algorithm get results in the form of an association rule, one of which is the connection of Oily Skin Clean Wash items -> Eshter Acne Toner with the highest support value of 1.0000. So it is expected that this can be used as a reference in efficient product sales going forward

Keywords : Data Mining, Apriori, Market Basket

1. INTRODUCTION

Along with the development of the era, more and more traders have also used information technology in their shops, both automated and manual systems for collecting data on sales transactions. The number of transactions that occur in one day is not small. These data are often only used as a recapitulation of goods input and inventory so that this data is often just left to accumulate like that. So the data processing is needed so that the existing data can be utilized for the benefit of business owners.

The process of data processing by extracting patterns from the data to get an added value which in this case is information in which the information cannot be known directly manually. [1] Data mining is useful to assist business owners in making decisions based on information obtained to assist the development of their business. [2] [10].

One algorithm that can be used is a priori algorithm which is used to obtain associations between items contained in a database. Usually this algorithm is used to find out whether an item has a correlation with other items. [8] [9] This is very suitable for business owners who want an efficient product sales strategy because by using this algorithm the business owner can find out which products might be bought as a set of items with other items. So business owners can sell products efficiently based on the possibilities previously obtained from calculations using the a priori algorithm so that business owners can not only sell one product but also other products based on the relationship between items or items. [11] [12]

The scope of this research is based on research on a priori algorithms that have been done by several previous researchers including research conducted by Dewi Kartika Pane entitled Implementation of Data Mining in Electronic Product Sales with Apriori Algorithms (Case Study: Credit Plus) [7] which examines the relationship of data sales of electronic goods using a priori algorithm Research conducted by Heru Dewantara, Purnomo Budi Santosa, Nasir Widha Setyanto entitled "Designing Data Mining Applications with Priori Algorithms for

Frequency of Shopping Cart Analysis on Sales Transaction Data" [12] where the research was used Universitas Brawijaya KPRI self-catering sales data.

The results of this research are the discovery of relationships between items that can produce information to increase product sales and this is what underlies this research is to determine product sales priorities.

2. RESEARCH METHOD

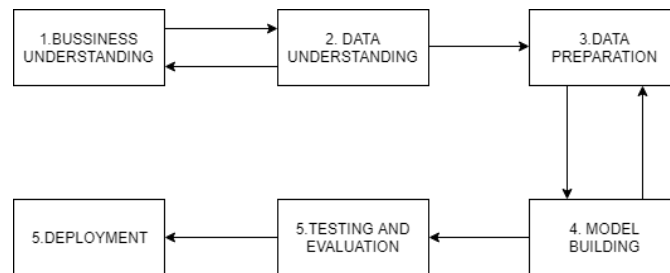


Figure 1. Steps of research with CRISP-DM method [1]

- a. **Business Understanding**
The business objective of Eshter House of Beauty is to sell beauty products. Eshter House of Beauty's need is to find patterns of consumer transactions to determine associations between products so that the owner can offer products with a package system.
- b. **Data Understanding**
At this stage, the data collection that will be used is the transaction data at the Eshter House of Beauty from September to October 2018
- c. **Data Preparation**
At this stage there is a data processing process namely transaction data in the Eshter House of Beauty from September to October 2018 as many as 277 transaction data which will be used to build a dataset that will be used in the modeling process.
- d. **Modeling**
In this modeling phase the association rules method will be used in the final data set, so that we can find out the associations between products in the data set.
- e. **Evaluation**
In this evaluation phase an evaluation will be carried out whether the method has achieved the objectives set at the beginning.
- f. **Deployment**
In this stage the results of data processing will be presented in the form of reports and graphs.

Apriori Algorithm

A priori algorithm is a method for finding patterns of relationships between one or more items in a dataset. [1] A priori algorithm is commonly used in transaction data or commonly called market basket, in the basketball market with an a priori algorithm, the owner of transaction data can find out a customer's buying patterns and what is the percent chance of a buyer to buy a shopping item simultaneously with other items.

A priori algorithm is a data collection algorithm with an associative rule (Association rule) to determine the associative relationship of a combination of items. Association Rule in question is done through the mechanism of calculating the support and confidence of an item relationship. An association rule is said to be interesting if the value of support is greater than

the minimum support and also the value of confidence is greater than the minimum confidence. [3]

The two main processes carried out in the Apriori algorithm, namely:

1. Join (merging).
In this process each item is combined with other items until no more combinations are formed.
2. Prune (pruning)
In this process, the results of the combined items are then trimmed using the minimum support specified by the user.

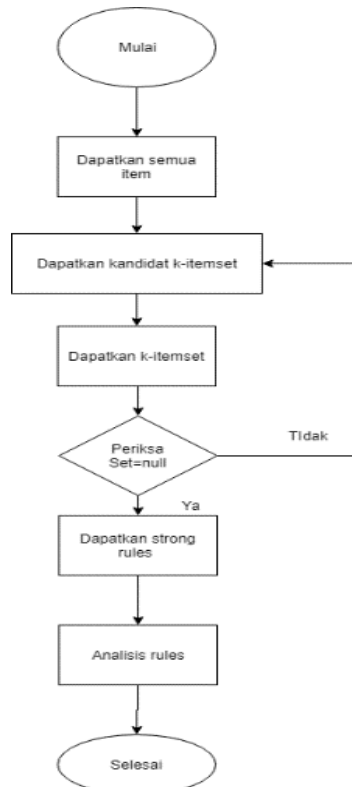


Figure 2. A priori Algorithm Flowchart [5]

Association Rule Method

Association rule is an analysis that aims to find out what relationships are owned between products purchased by consumers. In the basketball market analysis to test the validity of the data required a very large amount of data so that, relations and correlations between goods can be known more clearly. [4] [5]

The basic methodology of the Association Rule is divided into two stages, as follows:

- a. High frequency pattern analysis This stage looks for combinations of items that meet the minimum requirements of the support value in the database. An item's support value is obtained by the following formula:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi yang mengandung A}}{\text{Jumlah Total Transaksi}} \quad (1)$$

The support formula explains that the support value is obtained by dividing the number of transactions containing item A (one item) by the total number of all transactions. Meanwhile, to find the support value of 2 items using the following formula:

$$\text{Support}(A,B) = P(A \cap B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah Total Transaksi}} \quad (2)$$

- b. Formation of associative rules After all high frequency patterns have been found, then the associative rules are sought that meet the minimum requirements for confidence by calculating the confidence of associative rules $A \rightarrow B$. The confidence values of the rules $A \rightarrow B$ are obtained from the following formula.

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah Total Transaksi}} \quad (3)$$

Application of Data Mining

The data that has been obtained and collected will then be entered into the analysis phase where this data will be processed using data mining with a priori algorithm using the association rule methodology. Association Rule is one way to find rules in a certain combination of items in data mining.

Association Rule has two parameters, including confidence and support. Confidence (value of certainty), namely the strong relationship between items in the associative and support rules (supporting value), namely the percentage of the combination of items in the database. [6] [7]. Steps in data mining:

1. Data Selection. The data used in this study are transaction data from Eshter Beauty House during the period 02/09 / 2018-30 / 10/2018 or sales data for 2 months
2. Data Preprocessing / Data Cleansing, This stage is the stage where the data that has been obtained is cleaned so that the data becomes consistent and no data is found that is blank and has duplicated data.
3. Coding Transformation, After passing the data cleaning stage, the data will be entered into the transformation process where the data will be transformed as needed.

Data mining

Data mining process is the process of processing data using an algorithm.

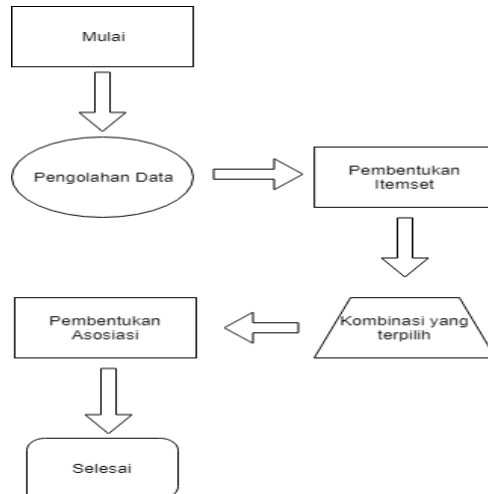


Figure 3. Data mining steps [6]

3. RESULT AND DISCUSSION

The dataset which previously consisted of Ref Date, Ref number, WNO, Item Code, Item Name, CAT NO, Sub Total was then processed into 3 columns according to the association rule processing requirements, namely date, Transaction Id, and item description.

Tanggal	Ref	Transaction.ID	Nama.Item
2/9/2018	1809020003	1809020003	ESTHER EYE CREAM (S)
2/9/2018	1809020003	1809020003	ESTHER GLOWING- C LOTION
4/9/2018	1809040001	1809040001	ESA LIPSTIK SWEET CORAL
4/9/2018	1809040001	1809040001	ESTHER HERBAL BODY SCRUB
4/9/2018	1809040001	1809040001	ESTHER BODY SCRUB
4/9/2018	1809040001	1809040001	ESA LIPSTIK FLURTY FUCHSIA
4/9/2018	1809040001	1809040001	ESA LIPSTIK WINE DIVA
5/9/2018	1809050002	1809050002	HEALTHY SKIN C (184)
5/9/2018	1809050002	1809050002	EHOB FACIAL SOAP
5/9/2018	1809050002	1809050002	EHOB TONER ALL SKIN
5/9/2018	1809050002	1809050002	ESTHER EYE CREAM (S)
5/9/2018	1809050002	1809050002	ESTHER SLIMMING CREAM
5/9/2018	1809050002	1809050002	HEALTHY SKIN A CREAM
5/9/2018	1809050002	1809050002	HEALTHY SKIN B (183)
6/9/2018	1809060005	1809060005	EHOB FACIAL SOAP
7/9/2018	1809070002	1809070002	ESTHER ACNE TONER
7/9/2018	1809070002	1809070002	ACNE GEL PLUS
7/9/2018	1809070002	1809070002	ESA LIPSTIK GLAM VIOLET
7/9/2018	1809070003	1809070003	ACNE CREAM PLUS (S)
7/9/2018	1809070003	1809070003	ESTHER ACNE TONER
7/9/2018	1809070002	1809070002	ESTHER ACNE CREAM
7/9/2018	1809070002	1809070002	CLEAN WASH OILY SKIN
8/9/2018	1809080007	1809080007	SUNBLOCK SPF 60
8/9/2018	1809080002	1809080002	ESTHER DAY CREAM (INTENSIF)
8/9/2018	1809080002	1809080002	ESTHER NIGHT CREAM (INTENSIF)
8/9/2018	1809080002	1809080002	ESTHER RADIANT TONER
8/9/2018	1809080002	1809080002	ESTHER GLOWING- C LOTION
8/9/2018	1809080002	1809080002	EHOB MILK CLEANSER

Figure 4. Pre-Process Data

After cleaning, the data will be sorted by transaction id or transaction number, which will be used to make the table according to the format in the association rule.

Table 1. Format of a priori data table

TID	Items
200	X Y Z
300	P Q Y
400	X Y P
500	C X Y

Previously random data is then sorted by transaction id contained in the dataset.

Tanggal.Ref	Transaction.ID	Nama.Item
1 2/9/2018	1809020003	ESTHER EYE CREAM (S)
2 2/9/2018	1809020003	ESTHER GLOWING- C LOTION
3 4/9/2018	1809040001	ESA LIPSTIK SWEET CORAL
4 4/9/2018	1809040001	ESTHER HERBAL BODY SCRUB
5 4/9/2018	1809040001	ESTHER BODY SCRUB
6 4/9/2018	1809040001	ESA LIPSTIK FLURTY FUCHSIA
7 4/9/2018	1809040001	ESA LIPSTIK WINE DIVA
8 5/9/2018	1809050002	HEALTHY SKIN C (184)
9 5/9/2018	1809050002	EHOB FACIAL SOAP
10 5/9/2018	1809050002	EHOB TONER ALL SKIN
11 5/9/2018	1809050002	ESTHER EYE CREAM (S)
13 5/9/2018	1809050002	HEALTHY SKIN A CREAM
14 5/9/2018	1809050002	HEALTHY SKIN B (183)
12 5/9/2018	1809050003	ESTHER SLIMMING CREAM
15 6/9/2018	1809060005	EHOB FACIAL SOAP
16 7/9/2018	1809070002	ESTHER ACNE TONER
17 7/9/2018	1809070002	ACNE GEL PLUS
18 7/9/2018	1809070002	ESA LIPSTIK GLAM VIOLET
21 7/9/2018	1809070002	ESTHER ACNE CREAM
22 7/9/2018	1809070002	CLEAN WASH OILY SKIN
19 7/9/2018	1809070003	ACNE CREAM PLUS (S)
20 7/9/2018	1809070003	ESTHER ACNE TONER
24 8/9/2018	1809080002	ESTHER DAY CREAM (INTENSIF)
25 8/9/2018	1809080002	ESTHER NIGHT CREAM (INTENSIF)
26 8/9/2018	1809080002	ESTHER RADIANT TONER
27 8/9/2018	1809080002	ESTHER GLOWING- C LOTION
28 8/9/2018	1809080002	EHOB MILK CLEANSER
29 8/9/2018	1809080002	SUNBLOCK SPF 60

Figure 5. Data Sorting Results

The data which had been sorted was processed again until it became like the format in Figure 3. For further processing by removing the date and reference number or transaction id. The results can be seen drawn below

itemList
1 ESTHER EYE CREAM (S),ESTHER GLOWING- C LOTION
2 ESA LIPSTIK SWEET CORAL,ESTHER HERBAL BODY SCRUB,...
3 HEALTHY SKIN C (184),EHOB FACIAL SOAP,ESTHER TONER AL...
4 ESTHER SLIMMING CREAM
5 EHOB FACIAL SOAP
6 ESTHER ACNE TONER,ACNE GEL PLUS,ESA LIPSTIK GLAM V...
7 ACNE CREAM PLUS (S),ESTHER ACNE TONER
8 ESTHER DAY CREAM (INTENSIF),ESTHER NIGHT CREAM (IN...
9 ESTHER SLIMMING CREAM
10 REJUVENATION EXTRA NIGHT CREAM,CLEAN WASH OILY S...
11 ESA BLACK PURIFYING PEEL OFF MASK
12 SUNBLOCK SPF 60,PIGMENT CONTROL CREAM,PIGMENT ...
13 ESTHER SUNCREAM (S),LOOSE POWDER ESTHER NATURAL
14 ESA BLACK PURIFYING PEEL OFF MASK,ESTHER RADIANT T...
15 ESTHER HAIR MIST,ESTHER SUNCREAM (L),ESA BLACK PUR...
16 EHOB BODY WASH FRESH FLORAL (L),EHOB TONER ALL S...
17 EHOB WHITENING BODY LOTION GREEN TEA (S),ESA SUN...
18 ESTHER FACE MIST (S),ESTHER HAIR MIST,EHOB BODY WA...
19 EHOB BODY BUTTER GREEN TEA (L),ESTHER ACNE LOTION...
20 ESA SUNNY PROTECTION CREAM,ESTHER PEELING CREAM...
21 EHOB MILK CLEANSER,ESTHER SUNCREAM (L),EHOB TON...
22 SUNBLOCK SPF 60
23 ESTHER CLEAN WASH
24 ESTHER EYE CREAM (L),ESTHER SUNCREAM (L)
25 ESTHER BRIGHTENING SOAP
26 ESTHER TEA TREE SERUM,ESA EGG CELLENT PEEL OFF MASK
27 ESTHER BRIGHTENING SOAP,REJUVENATION DAY NIGHT C...
28 ESTHER RADIANT TONER,ESTHER BRIGHTENING SOAP,SU...

Figure 6. Formation of Itemset

After processing, the data will be displayed in graphic form in the form of the most frequently purchased data and its correlation to other items

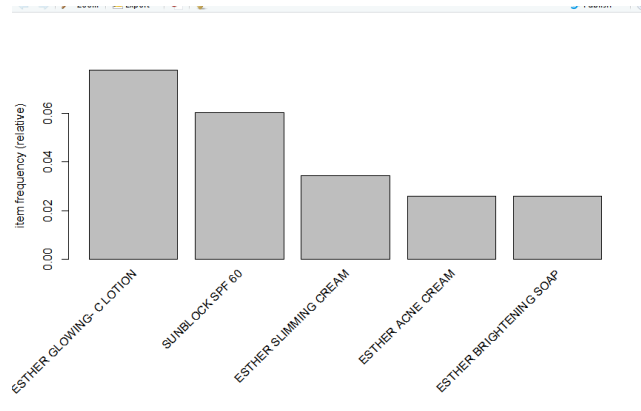


Figure 7. Frequent Items

Figure 7 shows the 5 items that appear most frequently in transaction data sets, which means the graph shows the items that appear most often in transaction data which can indicate that the item is the most salable item and then the diagram can be used as a reference for the main product in determining Companion products in the rule that will be sought later.

Then the data is processed using the association rules method to produce results in the form of rules. This rule is a picture of association between items where there are variables in the form of support, confidence and lift.

rules	support	confidence	lift	count
1 {ESTHER DAY CREAM (INTENSIF)} => {ESTHER NIGHT CREAM (INTENSIF)}	0.01626016	0.6666667	27.333333	2
2 {ESTHER NIGHT CREAM (INTENSIF)} => {ESTHER DAY CREAM (INTENSIF)}	0.01626016	0.6666667	27.333333	2
3 {ESTHER DAY CREAM (INTENSIF)} => {SUNBLOCK SPF 60}	0.01626016	0.6666667	4.100000	2
4 {ESTHER NECK CREAM} => {EHOB FACIAL SOAP}	0.01626016	1.0000000	11.181818	2
5 {PIGMENT CONTROL CREAM} => {PIGMENT CONTROL SERUM}	0.02439024	1.0000000	30.750000	3
6 {PIGMENT CONTROL SERUM} => {PIGMENT CONTROL CREAM}	0.02439024	0.7500000	30.750000	3
7 {PIGMENT CONTROL CREAM} => {ESTHER RADIANT TONER}	0.01626016	0.6666667	9.111111	2
8 {PIGMENT CONTROL CREAM} => {SUNBLOCK SPF 60}	0.01626016	0.6666667	4.100000	2
9 {PIGMENT CONTROL CREAM} => {ESTHER BRIGHTENING SOAP}	0.01626016	0.6666667	4.315789	2
10 {LOOSE POWDER ESTHER NATURAL} => {ESTHER SUNCREAM (S)}	0.01626016	0.6666667	6.833333	2
11 {HEALTHY SKIN B (183)} => {HEALTHY SKIN C (184)}	0.01626016	0.6666667	27.333333	2
12 {HEALTHY SKIN C (184)} => {HEALTHY SKIN B (183)}	0.01626016	0.6666667	27.333333	2
13 {HEALTHY SKIN B (183)} => {HEALTHY SKIN A CREAM}	0.01626016	0.6666667	27.333333	2
14 {HEALTHY SKIN A CREAM} => {HEALTHY SKIN B (183)}	0.01626016	0.6666667	27.333333	2
15 {HEALTHY SKIN B (183)} => {SUNBLOCK SPF 60}	0.01626016	0.6666667	4.100000	2
16 {HEALTHY SKIN C (184)} => {HEALTHY SKIN A CREAM}	0.01626016	0.6666667	27.333333	2
17 {HEALTHY SKIN A CREAM} => {HEALTHY SKIN C (184)}	0.01626016	0.6666667	27.333333	2
18 {CLEAN WASH OILY SKIN} => {ESTHER ACNE TONER}	0.03252033	1.0000000	12.300000	4
19 {PIGMENT CONTROL SERUM} => {ESTHER RADIANT TONER}	0.01626016	0.5000000	6.833333	2
20 {PIGMENT CONTROL SERUM} => {SUNBLOCK SPF 60}	0.01626016	0.5000000	3.075000	2

Figure 8. Rules generated based on association rules

	lhs	rhs	support	confidence	lift	count
[1]	{ESTHER DAY CREAM (INTENSIF)}	=> {ESTHER NIGHT CREAM (INTENSIF)}	0.01626016	0.6666667	27.333333	2
[2]	{ESTHER NIGHT CREAM (INTENSIF)}	=> {ESTHER DAY CREAM (INTENSIF)}	0.01626016	0.6666667	27.333333	2
[3]	{ESTHER DAY CREAM (INTENSIF)}	=> {SUNBLOCK SPF 60}	0.01626016	0.6666667	4.100000	2
[4]	{ESTHER NECK CREAM}	=> {EHOB FACIAL SOAP}	0.01626016	1.0000000	11.181818	2
[5]	{PIGMENT CONTROL CREAM}	=> {PIGMENT CONTROL SERUM}	0.02439024	1.0000000	30.750000	3
[6]	{PIGMENT CONTROL SERUM}	=> {PIGMENT CONTROL CREAM}	0.02439024	0.7500000	30.750000	3
[7]	{PIGMENT CONTROL CREAM}	=> {ESTHER RADIANT TONER}	0.01626016	0.6666667	9.111111	2
[8]	{PIGMENT CONTROL CREAM}	=> {SUNBLOCK SPF 60}	0.01626016	0.6666667	4.100000	2
[9]	{PIGMENT CONTROL CREAM}	=> {ESTHER BRIGHTENING SOAP}	0.01626016	0.6666667	4.315789	2
[10]	{LOOSE POWDER ESTHER NATURAL}	=> {ESTHER SUNCREAM (S)}	0.01626016	0.6666667	6.833333	2

Figure 9. Sample rules

The above rules are obtained from the test results with parameters namely support 0.01 or confidence 0.5, then 128 rules are generated. The rules generated above have the attributes in the form of, lift as the strength level of rules, support is the frequency of items contained in the rules in combination, while confidence is the value of strength between items in one rule. So, from some of the rules generated from the above test it is known that if consumers buy Eshter Day Cream (Intensive) then there is a possibility that customers buy Eshter Night Cream (intensive) with certainty of 0.6666667 and based on the lift value or the strength value of the rules of 27.3333 then the more it is likely that customers will buy the product.

4. CONCLUSION

Based on the results of research using a priori algorithm and testing with RStudio application, the conclusions can be drawn as follows:

1. Data Mining can be implemented on beauty product sales data by finding trends in combination patterns or relationships between items so that it can produce information that is useful in supporting decision making
2. The results obtained with a priori algorithm one of them, is a combination pattern that has the highest support value, namely Clean Wash Oily Skin -> Eshter Acne Toner with a support value of 1.00000
3. The more amount of data that is owned, the more accurate the values of the rules produced.

5. SUGGESTIONS

In data processing, it is better to use more data so that rules can be found more accurately in addition to the need for comparison with other algorithms in order to measure the level of accuracy that can be generated a priori algorithm.

REFERENCES

- [1] Ristianingrum, dan Sulastri, 2011, *"IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA APRIORI"*, Prosiding SINTAK 2017, pp. 372-382
- [2] D. Haryanto, Y. Oslan and D. Dwiwana, 2011, *Implementasi Analisa Keranjang Belanja Dengan Aturan Asosiasi Menggunakan Algoritma Apriori Pada Penjualan Suku Cadang Sepeda Motor*, Jurnal Buana Informatika, pp. 81-94.
- [3] Rinna Rachmatika, 2018, *Analisa Market Basket Menggunakan Algoritma Apriori Untuk STRATEGI PENJUALAN BERDASARKAN CUSTOMER FOCUS*, Prosiding Seminar Nasional Informatika, pp 1 – 6
- [4] Kennedy Tampubolon, Hoga Saragih, Bobby Reza, 2013, *IMPLEMENTASI DATA MINING ALGORITMA APRIORI PADA SISTEM PERSEDIAAN ALAT-ALAT KESEHATAN*, Informasi dan Teknologi Ilmiah (INTI), pp 93 -106
- [5] Nadya Rahmawati, Yuki Novia Nasution, Fidia Deny Tisna Amijaya *Aplikasi Data Mining Market Basket Analysis Untuk Menemukan Pola Pembelian di Toko Metro Utama BalikPapan*, Jurnal EKSPONENSIAL Vol. 8, No 1, Mei 2017, pp 1 -8
- [6] Moh.Sholik, Abu Salam, 2018, *Implementasi Algoritma Apriori untuk Menacarai Asosiasi Barang yang Dijual di E-commerce OrderMas*, Techno.COM, Vol. 17, No. 2
- [7] Dewi Kartika Pane, 2013, *IMPLEMENTASI DATA MINING PADA PENJUALAN PRODUK ELEKTRONIK DENGAN ALGORITMA APRIORI (STUDI KASUS : KREDITPLUS)*, Pelita Informatika Budi Darma, Volume : IV, Nomor: 3,

- [8] Robi Yanto, Riri Khoirah, 2015, *Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat*, *Citec Journal*, Vol. 2, No. 2,
- [9] Relita Buaton, Yani Maulita, Andri Kristiawan, 2018 *„Korelasi Faktor Penyebab Tindak Kekerasan dalam Rumah Tangga Menggunakan Data Mining Algoritma A Priori*, *Jurnal Media Infotama* Vol. 14 No. 1
- [10] Efori Buulolo, 2013, *IMPLEMENTASI ALGORITMA APRIORI PADA SISTEM PERSEDIAAN OBAT (STUDI KASUS : APOTIK RUMAH SAKIT ESTOMIHI MEDAN)*, *Pelita Informatika Budi Darma*, Volume : IV, Nomor: 1
- [11] Prof. Paresh Tanna, Dr. Yogesh Ghodasara 2013, *Foundation for Frequent Pattern Mining Algorithms' Implementation*, *International Journal of Computer Trends and Technology (IJCTT)* – Volume 4 Issue 7
- [12] Heru Dewantara, Purnomo Budi Santosa, Nasir Widha Setyanto, 2013, *Perancangan Aplikasi Data Mining Dengan Algoritma Priori Untuk Frekuensi Analisis Keranjang Belanja Pada Data Transaksi Penjualan*, *Jurnal Rekayasa dan Manajemen Sistem Industri* Vol 1, Nomor 3