published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

journal homepage: www.idss.iocspublisher.org/index.php/jidss

Application of Data Mining in Determining Sales Patterns at 212 Mart Supermarkets in Lubuk Pakam Using the Apriori Algorithm

Anzelia[1*], Penda Sudarto Hasugian[2]

[1,2] Informatics Engineering, STMIK Pelita Nusantara, Jl. Iskandar Muda No. 1 Medan

[1]anzelialia764@gmail.com*,[2]penda.hasugian@gmail.com
* Corresponding author

ARTICLE INFO

Article history:
Received: 19/08/2020
Revised: 23/10/2020
Accepted: 07/11/2020
Available online 01/12/2020

Keywords:
Apriori Algorithm,
Association Rule,
Data Mining,
Sales Pattern.

ABSTRACT

Data mining is a search and analysis on a very large database so that an interesting pattern is found with the aim of extracting information and knowledge that is accurate and potential, as well as understandable and useful from a large database. In this final project, one of the data mining techniques is implemented, namely the a priori algorithmcan define the rule pattern of the data set. The data in this study aims to determine the pattern of sales at 212 mart supermarkets in Lubuk Pakam. This research is obtained from the results of transaction data in the form of consumer purchase receipts, the data used are 10 transaction data with 43 total products. From the results of manual calculations, there are 23 rule association results in accordance with the minimum limit values for support and confidence that have been determined and the highest analysis results are found in french fries and chitato with a minimum support of 30% and 100% confidence. The results of the data mining process obtained can be used for the arrangement or arrangement of layout patterns that are adjusted to the association rules to suit consumer purchasing patterns.

1. Introduction

Currently the development of technology is growing rapidly in the world, especially in the world of the sales industry, the development and competition in the world of trade and advances in information technology is one thing that is related, in today's intense market competition, companies need the right strategy to be able to continue to meet market demands. . especially in supermarkets, 212 Mart is one of the supermarkets in Lubuk Pakam, 212 Mart is a Koperasi Syariah 212 minimarket, which sells people's daily necessities such as staples, household items, stationery, snacks etc. The problems encountered at 212 Mart are frequent empty goods, and the placement of product items that are not in accordance with consumer habits in buying goods simultaneously at one time. Determination of product layout is done to make it easier for consumers to find products simultaneously such as which products are suitable to be combined with other products, so as to save time for consumers, and as a decision aid in placing frequently purchased products together in an adjacent area. For this reason, it is necessary to have a method to determine sales patterns, namely using the Apriori Algorithm method.

By utilizing the a priori algorithm to study association rules, looking for patterns of relationships between one or more items in a dataset [5], the a priori algorithm is widely used in transaction data or commonly called the market basket, with the apriori algorithm the owner of a supermarket can find out the purchase pattern of a consumer. The previous research that the writer used as a reference related to this research using data mining method with a priori algorithm, namely: Research conducted by [2] with the title "Data mining analysis of product purchase patterns using the Apriori Algorithm method". Research conducted by [12] with the title "Implementation of data mining to analyze patterns of sales of goods using the Apriori Algorithm". The result of this research is that the a priori algorithm can analyze the product in the transaction so that it can be a reference for making decisions, the implementation is done by using the database as a temporary data storage combination.

2. Method

2.1 Research Framework

The following is the method or research framework carried out in this study which is described in the form of a flowchart.

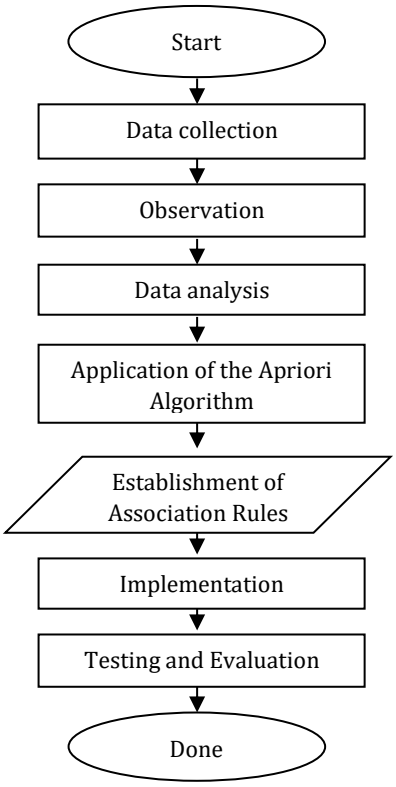


Figure 1. Research Framework

2.2 Data collection

Data collection is a very important stage in conducting research, at this stage an identification and collection of all data obtained from the 212 mart supermarket is carried out in order to obtain the data and information needed as supporting material which is indispensable in this study. In this study, data collection was carried out by means of observation.

Observation is a method of collecting data by observing directly in the field. In this study, observations were made directly at the 212 mart Lubuk Pakam supermarket. In this study using secondary data types, secondary data is a source of research data obtained through intermediary media or indirectly in the form of books, notes, existing evidence or archives both published and not generally published.

2.3 Method of Analysis with Apriori Algorithm

Association analysis is also known as one of the data mining techniques which is the basis of 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 efficient algorithms [6] . The associative rule is usually expressed in the following form:

If A then B [support = 2%, confidence = 60%], where A and B are a collection of items purchased by consumers of company X

This means: items A and B are purchased together by 2% of the overall transaction data analyzed and 60% of all consumers who buy item A also buy item B.

The basic methodological stages of association analysis are divided into two stages [11], namely:

1. High frequency pattern analysis

At this stage, look 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:

$$\text{Support (A)} = \frac{jumlahtransaksimengandungA}{totaltransaksi} \quad (1)$$

While the support value of 2 items is obtained from the following formula:

Support (A, B) = P (A∩B) (2)

$$\text{Support (A, B)} = \frac{\sum jumlahtransaksimengandungAdanB}{\sum totaltransaksi}$$

2. Establishing associative rules

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

$$\text{Confidence} = P(B \mid A) = \frac{\sum jumlahtransaksimengandungAdanB}{\sum jumlahtransaksimengandungA} \quad (3)$$

3. Results and Discussion

3.1. Data Analysis and Design Methods

Data analysis is a process carried out to convert data from research results into information that can later be used in drawing conclusions. The following is the initial data table that will be used as data mining calculations in this study:

Table 1
Preliminary data

Transaction Code	Item Name	The price of the item	Transaction date
10001001200600109	Bimoli Spc Pouch1l	16,000	01/04/2020
	Cap Tani Rice	7,900	
	French Fries	5,100	
	Chitato BEEF BBQ	9,500	
	Sarimi Duo K Chicken	6,400	
	Sdp Mie Soto 75gr	2,500	
	Fried Noodle Sedaap	2,600	
	Pepsodent	22,700	
	White sugar	7,500	
	Super Pell Cherry Ro	13,000	
	Telp Sbn Btgt	19,200	

The following is a table of transformed data. In the data transformation stage, the data to be transformed is adding numbers, changing the transaction code. Here are 10 transaction data that have been transformed, can be seen in table 2 below:

Table 2
Sales Transaction Data that has been transformed

No.	Transaction Code	Item Name
1	001	Bimoli spc, rice stamp tani, French fries, Chitato beef bbq, Sarimi duo k ayam, Sdp soto noodles, Sedaap mie goreng, Pepsodent, White sugar, Super pell cherry ro, Telp sbn btg
2	002	Cheetos, Lays rmpt lt, Walls fes van, French fries, Chitato beef bbq, Walls rainbow power, Aice choco cokies, Aice durian
3	003	French fries, Chitato beef bbq, Walls rainbow power, Cheetos
4	004	Cup 16, Stamped rice, White sugar, Flat cup lid, Coffee stamp si bridges, Indocafe cappuccino
5	005	Stamped rice, rice paper 1 bks
6	006	Giant prawn crackers, white sugar, stamped tani rice, purple potato chips, eggs

No.	Transaction Code	Item Name
7	007	Bimoli spc, Sunlight, Attack plus softener, Wipol carbol, Enzim clasic mild, Lif bw mc, Sbn btg Phone, Pesodent, Super pell cherry ro, Kispray, Faber castlel besar, Idm chicken broth, Sajiku versatile flour, 1 ounce palm sugar
8	008	Papaya mamaya soap, Wipol carbolic acid, Pepsodent, Wardah aloe fw, Samantha c.bath ging, Phone sbn btg, Super pell cherry ro
9	009	White sugar, Sariwangi, Unibis short cake, Sania pouch, small bbk martin coffee
10	010	Kopi Cap Jembatan Si, Cup 16, Paseo Smart, Sedaap Mie Soto, Sedaap Mie Goreng, Indocafe Cappuccino

Next, do the calculation of the data mining stage with a priori algorithm, namely by looking for the value of support and confidence in potential combinations or the formation of itemset for each item, by dividing the number of transactions for each of the total transactions multiplied by 100%. = Bimoli spc and so on.

Support (A) = $\frac{jumlahtransaksimengandungA}{totaltransaksi} * 100\,\%$

The data table can be seen below:

Table 3
Search for candidates for Support 1 itemset

No.	Item	amount	Candidate Support 1 itemset
1	Bimoli spc	2	Support (A1) $= \frac{2}{10} x100\% = 20\%$
2	Stamped farmer rice	4	Support (A2) $= \frac{4}{10} x100\% = 40\%$
3	French fries	3	Support (A3) $= \frac{3}{10} x100\% = 30\%$
4	Chitato beef bbq	3	Support (A4) $= \frac{3}{10} x100\% = 30\%$
5	Sarimi duo k chicken	1	Support (A5) $= \frac{1}{10} x100\% = 10\%$
....	.....	....	.....
43	Paseo smart	1	Support (A43) $= \frac{1}{10} x100\% = 10\%$

With the support value obtained, a minimum support of 20% can be determined with a minimum number of appearances = 2, then eliminating the support value of 1 itemset that does not meet the minimum support requirements, here is a table that meets 1 itemset support.

Table 4
Minimum result of support for 1 itemset

No.	Item Code	Item Name	amount	Results Support 1 itemset
1	A1	Bimoli spc	2	20%
2	A2	Stamped farmer rice	4	40%
3	A3	French fries	3	30%
4	A4	Chitato beef bbq	3	30%
5	A6	Sdp Mie Soto	2	20%
....	.....	.....	...	.....
16	A28	Carbolic Wipol	2	20%

Next, determine support for 2 itemset, from the transaction data in table 5 and the results of the minimum support for 1 itemset in table 4.6 above, by combining 2 item sets. then you can determine the support value for 2 itemset with the following formula:

Support (A, B) = $\frac{\sum jumlahtransaksimengandungAdanB}{totaltransaksi} * 100\%$

Table 5
Candidate support for 2 itemset

No.	Item	amount	Candidate support for 2 itemset
1	Bimoli spc, stamp tani rice	1	Support (A1, A2) $= \frac{1}{10} x100\% = 10\%$
2	Bimoli spc, French fries	1	Support (A1, A3) $= \frac{1}{10} x100\% = 10\%$

No.	Item	amount	Candidate support for 2 itemset
3	Bimoli spc, Chitato beef bbq	1	Support (A1, A4) $= \frac{1}{10} x100\% = 10\%$
4	Bimoli spc, Sdp mie soto	1	Support (A1, A6) $= \frac{1}{10} x100\% = 10\%$
5	Bimoli spc, Sedaap fried noodles	1	Support (A1, A7) $= \frac{1}{10} x100\% = 10\%$
....	.....	.....	.....
120	Indocafe cappuccino, Wipol carbol	0	Support (A21, A28) $= \frac{0}{10} x100\% = 0\%$

After getting a support value of 2 itemset, a minimum support of 20% can be determined, eliminating the 2 itemset support value that does not meet the minimum support requirements as follows:

Table 6
Minimum result of support for 2 itemset

No.	Item Code	Item Name	amount	*support* 2 itemset
1	(A1, A8)	Bimoli spc, Pepsodent	2	20%
2	(A1, A10)	Bimoli spc, Super pell cherry ro	2	20%
3	(A1, A11)	Bimoli spc, Telp sbn btg	2	20%
4	(A2, A9)	Stamped tani rice, white sugar	3	30%
5	(A3, A4)	French fries, Chitato beef bbq	3	30%
...	.....	.....	...	.....
20	(A20, A21)	Coffee stamp si bridges, Indocafe cappuccino	2	20%

After getting the minimum support for 2 itemset, the next step is to calculate the confidence value, the confidence value is determined from each combination in table 4.6, using the following formula:

$$\text{Confidence} = P(B \mid A) = \frac{\sum jumlahtransaksimengandungAdanB}{\sum jumlahtransaksimengandungA} * 100\%$$

Table 7
Looking for the value of confidence

No.	Item Name	Amount A	Total A∩B	*Confidence* %
1	Bimoli spc, Pepsodent	2	2	$\frac{2}{2} x100\% = 100\%$
2	Pepsodent, Bimoli spc	3	2	$\frac{2}{3} x100\% = 66{,}66\%$
3	Bimoli spc, Super pell cherry ro	2	2	$\frac{2}{2} x100\% = 100\%$
4	Super pell cherry ro, Bimoli spc	3	2	$\frac{2}{3} x100\% = 66{,}66\%$
5	Bimoli spc, Telp sbn btg	2	2	$\frac{2}{2} x100\% = 100\%$
...	.....	.....	.....	.....
28	Indocafe cappuccino, coffee stamp si bridges	2	2	$\frac{2}{2} x100\% = 100\%$

After getting the confidence value, then determine the minimum confidence value of 75%, and eliminate the confidence value that does not meet the minimum confidence requirement. then then create a table of association rules that has been formed as follows:

Table 8
Formed Association Rules Table

No.	Item Name	*Support*	*Confidence*
1	Bimoli spc, Pepsodent	20%	100%
2	Bimoli spc, Super pell cherry ro	20%	100%
3	Bimoli spc, Telp sbn btg	20%	100%
4	Stamped tani rice, white sugar	40%	75%
5	White sugar, rice with peasant stamp	40%	75%
6	French fries, Chitato beef bbq	30%	100%
7	Chitato beef bbq, French fries	30%	100%
....	.....	.....	.....

No.	Item Name	*Support*	*Confidence*
23	Indocafe cappuccino, coffee stamp si bridges	20%	100%

Based on the results of the confidence calculations that have been formed above, from the table it is obtained the confidence value where all confidence values have met the minimum value, it can be concluded that the combination rule most often purchased by consumers is French fries and Chitato beef bbq with 30% support and 100% confidence. , here is a description of the rules that have been obtained, as many as 23 rule items.

Rule 1 : If consumers buy French fries, they will buy chitato beef bbq with 30% support and 100.00% confidence.

Rule 2 : If consumers buy chitato beef bbq, they will buy French fries with 30.00% support and 100.00% confidence.

Rule 3 : If consumers buy Sedaap Mie Soto, they will buy Sedaap fried noodles with 20% support and 100.00% confidence.

Rule 4 : If consumers buy Sedaap fried noodles, they will buy Sedaap Soto noodles with 20% support and 100.00% confidence.

Etc .....

Rule 23: If consumers buy white sugar, they will buy farmer-stamped rice with 30.00% support and 75.00% confidence.

3.2 System Implementation

a. *Use Case* Diagram

Use Case The diagram is a usecase diagram that is used to briefly describe who uses the system and what it does, here is a usecase diagram based on this research.

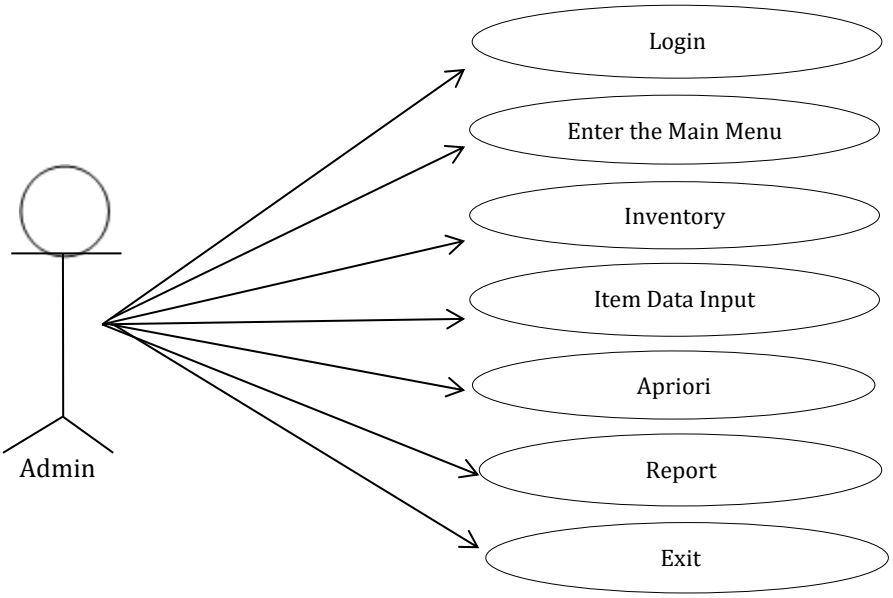


Figure 2. Use Case Diagram

The following is a description of the Use Case Diagram that has been made.

Table 9
Description of the Use Case Diagram

Use Case Name	Description
Login	The first step Admin must log in
Enter the main menu	After logging in the system will display the main menu
Inventory	Admin can input incoming goods and goods sold
Item Data Input	Admin can input incoming goods data and sold goods data
A priori provisions	Admin can determine the value of the minimum support and confidence conditions

Use Case Name	Description
Report	The report will display the results of a priori calculations and rule provisions and will also display monthly reports and tabulation tables
Exit	Admin Exit system

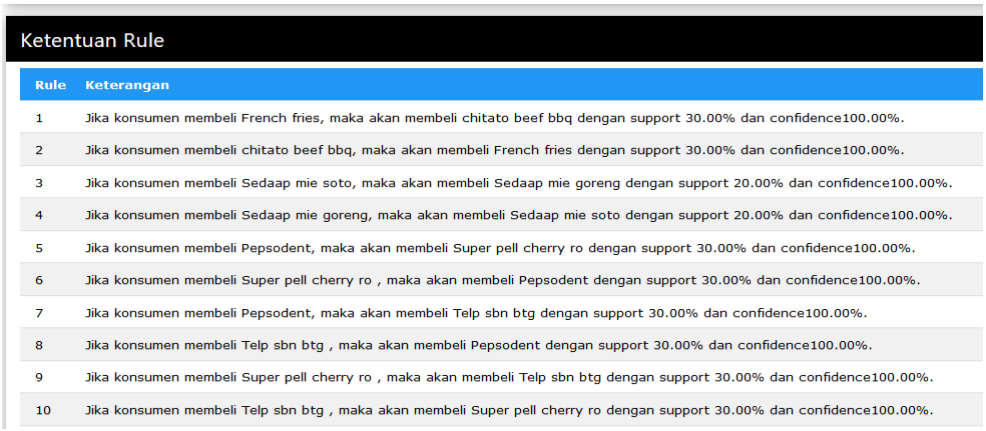
b. Result Display

The following is a display of confidence results and results in the form of rules from the system that has been built as follows:

No.	Kode Item	Nama	Frekuensi A	Frekuensi A dan B	Confidence
1	{ A1 } -> { A8 }	Bimoli spc, Pepsodent	2	2	100.00%
2	{ A1 } -> { A10 }	Bimoli spc, Super pell cherry ro	2	2	100.00%
3	{ A1 } -> { A11 }	Bimoli spc, Telp sbn btg	2	2	100.00%
4	{ A2 } -> { A9 }	Beras cap tani, Gula putih	4	3	75.00%
5	{ A9 } -> { A2 }	Gula putih , Beras cap tani	4	3	75.00%
6	{ A3 } -> { A4 }	French fries, chitato beef bbq	3	3	100.00%
7	{ A4 } -> { A3 }	chitato beef bbq, French fries	3	3	100.00%
8	{ A6 } -> { A7 }	Sedaap mie soto, Sedaap mie goreng	2	2	100.00%
9	{ A7 } -> { A6 }	Sedaap mie goreng, Sedaap mie soto	2	2	100.00%
10	{ A8 } -> { A10 }	Pepsodent, Super pell cherry ro	3	3	100.00%

Figure 3. *Confidence Results*

In the confidence results there is a display of the results of associations that have been formed and there are 23 rule combinations, on the display there are item codes, item names and the minimum value of support and confidence.



Rule	Keterangan
1	Jika konsumen membeli French fries, maka akan membeli chitato beef bbq dengan support 30.00% dan confidence100.00%.
2	Jika konsumen membeli chitato beef bbq, maka akan membeli French fries dengan support 30.00% dan confidence100.00%.
3	Jika konsumen membeli Sedaap mie soto, maka akan membeli Sedaap mie goreng dengan support 20.00% dan confidence100.00%.
4	Jika konsumen membeli Sedaap mie goreng, maka akan membeli Sedaap mie soto dengan support 20.00% dan confidence100.00%.
5	Jika konsumen membeli Pepsodent, maka akan membeli Super pell cherry ro dengan support 30.00% dan confidence100.00%.
6	Jika konsumen membeli Super pell cherry ro , maka akan membeli Pepsodent dengan support 30.00% dan confidence100.00%.
7	Jika konsumen membeli Pepsodent, maka akan membeli Telp sbn btg dengan support 30.00% dan confidence100.00%.
8	Jika konsumen membeli Telp sbn btg , maka akan membeli Pepsodent dengan support 30.00% dan confidence100.00%.
9	Jika konsumen membeli Super pell cherry ro , maka akan membeli Telp sbn btg dengan support 30.00% dan confidence100.00%.
10	Jika konsumen membeli Telp sbn btg , maka akan membeli Super pell cherry ro dengan support 30.00% dan confidence100.00%.

Figure 4. Results of Rule Provisions

In the results of the rule provisions, there are descriptions of item combination patterns in the form of text from the confidence results above where there are 23 rules and the highest value is in French fries with 30% support and 40% confidence.

4. Conclusion

Based on the research and implementation that has been done in Application of Data Mining To determine sales patterns at 212 mart supermarkets Using the Apriori Algorithm, then the conclusions that can be drawn from this research are as follows:

a. A priori algorithm can be implemented to find sales patterns using sales transaction data.
b. The research was conducted by searching for association rules through processing data on purchases of goods from consumers, then looking for the relationship between the items purchased so that the data obtained will produce an association rule so that it can be used for proper layout settings and can improve sales strategies.
c. In analyzing a number of data, that the smaller the minimum value of support and confidence is determined, the more rules will be generated, and if the bigger the minimum limit of support and confidence is determined, the fewer rules will be generated.
d. This system is able to process transaction data to find frequent itemset and association rules that meet transaction limits and are able to display rules in text form.

5. Reference

[1] Friday Kurniasih, Nia Kumaladewi, Bakri La Katjong, "Analisa Dan Perancangan Data Minng Dengan Metode Market Basket Analysis Untuk Analisa Pola Belanja Konsumen Pada Tendecies Store", Jurnal Sistem Informasi, 5(1), 2012, 1-10.
[2] Heroe Santoso, I Putu Hariyadi, Prayitno, "Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori", Seminar Nasional Teknologi Informasi Dan Multimedia 2016, ISSN : 2302-3805.
[3] Irsyad Djamaludin, "Analisis Pola Pembelian Konsumen Pada Transaksi Penjualan Menggunakan Algoritma Apriori", Jurnal SIMETRIS, Vol.8 No.2, November 2017, 2252-4983.
[4] Kadek Juni Arta, Gede Indrawan, Gede Rasben Dantes, "Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di STMIK Denpasar Menggunakan Metode *Technique For Others Reference By Similarity To Ideal Solution*", Vol.5, No.2, Oktober 2016.
[5] Kristiani Romili Sitanggang, Penda Sudarto Hasugian, "Penerapan Data Mining Dalam Menganalisa Pola Peminjaman Buku di Perpustakaan SMP Negeri 2 Beringin Satu Atap Menggunakan Algoritma Apriori", JIKOMSI Jurnal Ilmu Komputer dan Sistem Informasi Vol.1No 1, Maret2018, pp 1-5.
[6] Mateus Paga Tana, Fitri Marisa, Indra Dharma Wijaya, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Pada Toko Oase Menggunakan Algoritma Apriori", Jurnal Informasi Merdeka Pasuruan. Vol.3, No.2, Agustus 2018.
[7] Muhammad Syahril, Kamil Erwansyah, Milfa Yetri, "Penerapan Data Mining untuk menentukan pola penjualan peralatan sekolah pada brand wigglo dengan menggunakan algoritma apriori", Jurnal teknologi sistem informasi dan sistem komputer. Vol.3, No.1, 118-136, Januari 2020.
[8] Nurdin, Dewi Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe", Techsi Vol.6, No.1, April 2015.
[9] Oktaviani Manurung, Penda Sudarto Hasugian, *"Analisa Algoritma Apriori Untuk Peminjaman Buku Pada Perpustakaan SMA 1 Silima Pungga-Pungga Parongil"*, Riset dan E-Jurnal Manajemen Informatika KomputerVolume 4, Number 1, Oktober 2019.
https://www.jurnal.polgan.ac.id/index.php/remik/article/view/10445
[10] Penda Sudarto Hasugian, Suprianto Panjaitan, "*THE DATA MINING OF CELL PHONE MOST INTERESTED USING APRIORIAL ALGORITHM",* JURNAL INFOKUM, Volume 7, No.1, Desember 2018ISSN 2302-9706.
http://infor.seaninstitute.org/index.php/infokum/index
[11] Sheih Al Syahdan, Anita Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota".Jurnal Nasional Komputasi dan Teknologi Informasi Vol. 1 No. 2, Oktober 2018 P-ISSN 2620-8342 E-ISSN 2621-3052.
https://scholar.google.co.id/citations?user=soOqtd0AAAAJ&hl=en
[12] Tony Suryaudin, "Implementasi Data Mining Untuk Menganalisa Pola Penjualan Barang Dengan Menggunakan Algoritma Apriori", Simki-Techsain, Vol.02, No.04, Tahun 2018, ISSN : 2599-3011.