Indexed by

Scopus

DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS

MARKET BASKET ANALYSIS OF ADMINISTRATIVE PATTERNS DATA OF CONSUMER PURCHASES USING DATA MINING TECHNOLOGY

Crossref

Lukman Samboteng
Politeknik STIA LAN Makassar, Makassar, Indonesia

Rulinawaty
Universitas Terbuka, Indonesia

M.Rachmat Kasmad
Universitas Negeri Makassar, Makassar, Indonesia

ROAD
DIRECTORY
OF OPEN ACCESS
SCHOLARLY
RESOURCES

Mutmainnah Basit
Universitas Negeri Makassar, Makassar, Indonesia

Robbi Rahim
Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia

KoBSON

SCINDEKS
Srpski citatni indeks

Google
Scholar

Cite article:

Online aceess of full paper is available at: *www.engineeringscience.rs/browse-issues*

MARKET BASKET ANALYSIS OF ADMINISTRATIVE PATTERNS DATA OF CONSUMER PURCHASES USING DATA MINING TECHNOLOGY

Lukman Samboteng[1,], Rulinawaty[2], M.Rachmat Kasmad[3], Mutmainnah Basit[3], Robbi Rahim[4]*
[1]Politeknik STIA LAN Makassar, Makassar, Indonesia
[2]Universitas Terbuka, Indonesia
[3]Universitas Negeri Makassar, Makassar, Indonesia
[4]Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia

Food is the ingredient that enables people to grow, develop, and achieve. For this reason, food quality and types of food must be considered so that they are safe for consumption and managed. Some plant-based foodstuffs are often processed and consumed by the community, even the most needed in food processing. In this case, the research was carried out using data mining with market basket analysis algorithms to obtain very valuable information to decide the inventory of the type of material needed. Market Based Analysis method is used to analyze all data and create patterns for each data. One method of Market Based Analysis in question is the association rule with a priori algorithm. This algorithm produces sales transactions with strong associations between items in the transaction which are used as sales recommendations that help users (owners) get recommendations when users see details of the itemset purchased. From the results of the trials in this study, it was found that the greater the minimum support (minsup) and minimum confidence (minconf), the less time it takes to produce recommendations and the fewer recommendations are given, but the recommendations given come from transactions that often appear.

Key words: apriori algorithms, transaction data, data mining, associations, consumer patterns

INTRODUCTION

There are certainly very important changes in the conditions in a company, in the environment and outside of the company. The fast and correct taking of business decisions is one solution in order for the company to grow further and adapt to existing changes [1]. Data management for sales transactions is one thing that can be used for decision making. Most transaction data will not be reused, stored only as an archive and used only to produce a selling report. To increase sales, a supermarket needs important information to increase its sales, the accumulated transaction data must be processed as useful data. These useful information can be used to develop a corporate policy, namely to implement the information in a corporate approach [2]. Supermarkets in their operation sell the public different kinds of fundamental needs in which sales transaction data from their sales operations exist. The data are used only as a reference and only as an archive for sales reports. Many computer branches, including data mining [3] can solve complex problems [4]. Data mining is a process that uses statistical, mathematical, artificial and machine learning techniques to obtain and identify useful knowledge and information from different large databases [5]. Data mining is an algorithm that is used for the treatment of data to find hidden patterns from the data processed. The data processed then produces new knowledge from old data, which can be used in the determination of future decisions [6]. Furthermore, the application of data mining can perform data analysis of item sales transaction data, which produces several association rules in customer purchasing patterns in supermarkets, which is why it is being implemented [7–9]. The process of searching for hidden patterns of data for each transaction whose previously unknown type of data is stored in databases, data warehouses, or other information storage media is referred to as data mining [10–13]. Several data mining methods are frequently used, including the Apriori Algorithm, which determines the itemset of book loan transactions, which are then grouped into itemset1 table, itemset2 table, support value table, confidence value table, and customer purchasing patterns. Many associated research uses Apriori's data mining branch. Research is one of them [14]. The results of this study show that 142 Association Rules meet the support of more than 10% and 50% confidence, based on sample data, can be applied as a whole. Based on test results using Rapidminer, the product that consumers want is a minimum confidence above 50 percent can be concluded. Based on the test results In addition, the author uses it a priori as easy to operate, the process needed is relatively short, flexible and user-friendly, and uses simple non-statistical principles.

RESEARCH METHOD

The investigation was done by collecting data on the basis of direct observation on the site in order to clearly identify existing problems. A survey is then conducted to obtain the necessary information or data. Use

historical data to find regularities for activities such as the collection. It is hoped that the problem can be understood by analysizing the predetermined problems. The decision-making process is described in the analysis of this problem in the administrative data on the purchase pattern of consumers. Using predetermined criteria, the analytical method used to analyze needs in optimal decisions was determined. The results of the research are then processed using the Apriori algorithm. The process stage in which the data processed is determined by frequently searching for items. In the following section the techniques employed for searching Frequent Sets and rules of association are analyzed and understood. RapidMiner is a solution for analyzing data mining, text mining and predictive analysis. RapidMiner uses a variety of descriptive and predictive techniques to provide users with insights so they can make the best decisions.

Data Mining

Data mining is the process of obtaining data from databases using algorithms and techniques that include statistics, mechanical study and databases, in which data extraction in a large variety of databases gives you a joy in the form of knowledge [11, 15–17]. Important matters concerning the mining of data are:

a. Automatic mining of existing data is an automatic process.

b. In the form of large data, the data which must be processed.

c. The purpose of data mining is to identify relationships or patterns that can be useful.

Association Rules

Databases on customer transactions in supermarkets or in other areas have been made available to facilitate the development of methods that find product groups or items automatically in the database [18]. An example is the transaction data for the supermarket. The transaction data lists all items that a customer purchases in a single buying transaction. The seller wants to know whether the customer always purchases the product together. Sellers may use this information to establish a supermarket layout to allow for optimal arrangements of the items for promotional purposes, the segmentation of buyers, and the associations to provide information by means of a 'if - then' relationship calculated on the basis of probabilistic evidence [19]. The idea of an association rule is to examine all possible connections between items and to select only those which will most likely be indicators of dependence [20]. The term predecent is usually used to indicate the "IF," which is the result of which is "THEN." Antendecent is a group of items which have no common relationship in this analysis [18]. The association rules are developed in two stages, the following [21]:

a. Analyzing patterns of high frequency. This phase seeks a combination of items which meet the minimum support value requirements in the database. The rule "X => Y," (1) is supported by the likelihood that a transaction will concurrently involve attributes or sets of X and Y attributes. The mathematical equation is: support (X =>Y) = P (X alternatively y) (2) Info: X => Y = items that are shown together P (X oscillating Y) — probability of X and Y transactions divided by the total transaction number.

b. Association Rules Establishment. Once all high frequency patterns are detected, see the association rule, which meets the minimum confidence requirements by calculating whether the X then Y rule is confident. The mathematical confidence formula is: Trust (X => Y) = P (Y | X) (3) Info: X => Y = items simultaneously displayed P (Y|X) = probability of the number of X and Y transactions divided by the number of X-containing transactions.

Apriori Algorithm

The Apriori algorithm is a fundamental algorithm for the determination of frequency sets of the bolean regimes [21]. The main idea is to search for frequent item sets first (a set of items that meet minimum support). Delete itemset based on the predetermined minimum support level from the second transaction database. Next, create itemset association rules, which meet the minimum data base confidence value [22]. The steps in the a priori algorithm [21]:

a. High-frequency analysis (frequent item set) for item combinations that comply with the minimum support value requirements for transaction data..

$$Support\ (A) = \frac{the\ amount\ the\ transaction\ contains\ A}{Total\ transactions}\ X\ 100\% \quad (1)$$

By the following formula the supporting value for the two items is achieved.

$$Support\ (A \cap B) = \frac{the\ amount\ the\ transaction\ contains\ A\ and\ B}{Total\ transactions}\ X\ 100\% \quad (2)$$

In general, the value of support can also be found using a formula:

$$Support\ = \frac{the\ number\ of\ items\ purchased}{the\ amount\ of\ all\ transactions}\ X\ 100\% \quad (3)$$

b. Establishing associative rules, to meet the minimum confidence requirements.

$$Confidence\ (A,B) = \frac{the\ amount\ the\ transaction\ contains\ A\ and\ B}{the\ amount\ the\ transaction\ contains\ A}\ X\ 100\% \quad (4)$$

In general, the confidence value can also be found using a formula:

Lukman Samboteng, et al. - Market basket analysis of administrative patterns data of consumer purchases using data mining technology

iipp

$$Confidence = \frac{the\ number\ of\ items\ purchased}{the\ number\ of\ transactions\ in\ the\ antecedent\ part} \times 100\% \quad (5)$$

RESULT AND DISCUSSION

This is an example of the data on day-to-day, basic sales transactions processed according to the research requirements of Tables 1 and 2, below.

Table 1: Data for Process Mining

Item Name	Number	Field
Chili	6	100
Onion	8	50
Tomato	6	30
Potato	5	40
Tempe	5	20

Table 2: Item Frequency

Transaction	Items Purchased				
	Chili	Onion	Tomato	Potato	Tempe
1	1	1	1	0	0
2	1	1	0	0	1
3	0	1	0	1	1
4	0	0	1	1	1
5	0	1	0	1	1
6	1	1	1	0	0
7	1	1	0	1	0
8	1	0	1	0	1
9	0	1	1	1	0
10	1	1	1	0	0
Total	6	8	6	5	5

Frequent itemset = 3 then the associations that may be formed are {Chili, Onion}, {Chili, Tomato}, {Chili, Potatoes}, {Chili, Tempe}, {Onions, Tomatoes}, {Onions, Potatoes}, {Onions, Tempeh }, {Tempe, Potatoes}, {Potatoes, Tempe}, {Tomatoes, Tempe}. From table 3 of the above elements, P means items sold concurrently, while S means that no item is sold concurrently or transactions occur. ·f is the frequency number of the set item. The number of the articles set shall be higher or equal to the number of set articles ($\sum > = \Phi$). The above table is obtained from: F2 = {{Chile, Onion}, {Chile, Tomato, Onion}}. The following are obtained from the following table: The combination of F2 elements can be combined into three candidates.

Table 3: Two Itemset

Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	1	P
2	1	1	P
3	0	1	S
4	0	0	S
5	0	1	S
6	1	1	P
7	1	1	P
8	1	0	S
9	0	1	S
10	1	1	P
Number of Transactions 2 Items			5
2 Support (%)			50

Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	1	P
2	1	0	S
3	0	0	S
4	0	1	S
5	0	0	S
6	1	1	P
7	1	0	S
8	1	1	P
9	0	1	S
10	1	1	P
Number of Transactions 2 Items			4
2 Support (%)			40

Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	0	S
2	1	0	S
3	0	1	S
4	0	1	S
5	0	1	S
6	1	0	S
7	1	1	P
8	1	0	S
9	0	1	S
10	1	0	S
Number of Transactions 2 Items			1
2 Support (%)			10

Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	0	S
2	1	1	P
3	0	1	S
4	0	1	S
5	0	1	S
6	1	0	S
7	1	0	S
8	1	1	P
9	0	0	S
10	1	0	S
Number of Transactions 2 Items			2
2 Support (%)			20
Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	1	P
2	1	0	S
3	1	0	S
4	0	1	S
5	1	0	S
6	1	1	P
7	1	0	S
8	0	1	S
9	1	1	P
10	1	1	P
Number of Transactions 2 Items			4
2 Support (%)			40
Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	0	S
2	1	0	S
3	1	1	P
4	0	1	S
5	1	1	P
6	1	0	S
7	1	1	P
8	0	0	S
9	1	1	P
10	1	0	S
Number of Transactions 2 Items			4
2 Support (%)			40

Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	0	S
2	1	1	P
3	1	1	P
4	0	1	S
5	1	1	P
6	1	0	S
7	1	0	S
8	0	1	S
9	1	0	S
10	1	0	S
Number of Transactions 2 Items			3
2 Support (%)			30
Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	0	S
2	0	0	S
3	0	1	S
4	1	1	P
5	0	1	S
6	1	0	S
7	0	1	S
8	1	0	S
9	1	1	P
10	1	0	S
Number of Transactions 2 Items			2
2 Support (%)			20
Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	0	0	S
2	0	1	S
3	1	1	P
4	1	1	P
5	1	1	P
6	0	0	S
7	1	0	S
8	0	1	S
9	1	0	S
10	0	0	S
Number of Transactions 2 Items			3
2 Support (%)			30

Lukman Samboteng, et al. - Market basket analysis of administrative patterns data of consumer purchases using data mining technology

iipp

Transaction	Items Purchased		2 item transaction
	Chili	Onion	
1	1	0	S
2	0	1	S
3	0	1	S
4	1	1	P
5	0	1	S
6	1	0	S
7	0	0	S
8	1	1	P
9	1	0	S
10	1	0	S
Number of Transactions 2 Items			2
2 Support (%)			20

Transaction	Items Purchased			3 item transaction
	Chili	Onion	Tomato	
1	1	0	0	S
2	0	0	1	S
3	0	1	1	S
4	1	1	1	P
5	0	1	1	S
6	1	0	0	S
7	0	1	0	S
8	1	0	1	S
9	1	1	0	S
10	1	0	0	S
Number of Transactions 3 Items				1
3 Support (%)				10

Table 4:Three Itemset

Transaction	Items Purchased			3 item transaction
	Chili	Onion	Tomato	
1	1	1	1	P
2	1	1	0	S
3	0	1	0	S
4	0	0	1	S
5	0	1	0	S
6	1	1	1	P
7	1	1	0	S
8	1	0	1	S
9	0	1	1	S
10	1	1	1	P
Number of Transactions 3 Items				3
3 Support (%)				30
Transaction	Chili	Onion	Tomato	3 item transaction
1	1	0	0	S
2	1	0	1	S
3	1	1	1	P
4	0	1	1	S
5	1	1	1	P
6	1	0	0	S
7	1	1	0	S
8	0	0	1	S
9	1	1	0	S
10	1	0	0	S
Number of Transactions 3 Items				2
3 Support (%)				20

Itemset - itemset that can be combined is an itemset - itemset that has the same value in the first k-1 item. For k = 3 (3 elements), the possible sets are: {Chili, Onion, Tomato}. From the tables above, we get F3 = {}, because there is no $\sum$> = Φ so that F4, F5, F6 and F7 are also empty sets. Determine (ss-s) as the antecendent and s as the cosequent of the obtained Fk. In F2, the set F2 = {{Chili, Onion} {Tomato Onion}, {Potato Onion}} is obtained. From the above steps, 10 rules are obtained, namely:

a.   If you buy Chili, it will buy Onions
b.   If you buy Chili, you will buy Tomatoes
c.   If you buy Chili, you will buy Potatoes
d.   If you buy Chili, you will buy Tempe
e.   If you buy Onions, you will buy Tomatoes
f.   If you buy Onions, you will buy Potatoes
g.   If you buy Onions, you will buy Tempe
h.   h. If you buy Tomatoes, you buy Potatoes
i.   i. If you buy Potatoes, you will buy Tempe
j.   j. If you buy Tomatoes, you will buy Tempe

Table 5: Multiplication of Support and Confidence

Rules	Support	Confidence	Support and Confidence
a. If you buy Chili, it will buy Onions	20%	30%	0,06
b. If you buy Chili, you will buy Tomatoes	16%	24%	0,0384
c. If you buy Chili, you will buy Potatoes	4%	6%	0,0024
d. If you buy Chili, you will buy Tempe	8%	12%	0,0096
e. If you buy Onions, you will buy Tomatoes	16%	24%	0,0384
f. If you buy Onions, you will buy Potatoes	16%	24%	0,0384
g. If you buy Onions, you will buy Tempe	12%	18%	0,0216
h. If you buy Tomatoes, you buy Potatoes	8%	12%	0,0096
i. If you buy Potatoes, you will buy Tempe	12%	18%	0,0216
j. If you buy Tomatoes, you will buy Tempe	8%	12%	0,0096

The following is a design model for the data entry process using the Rapid Miner software as shown in Figure 1 below
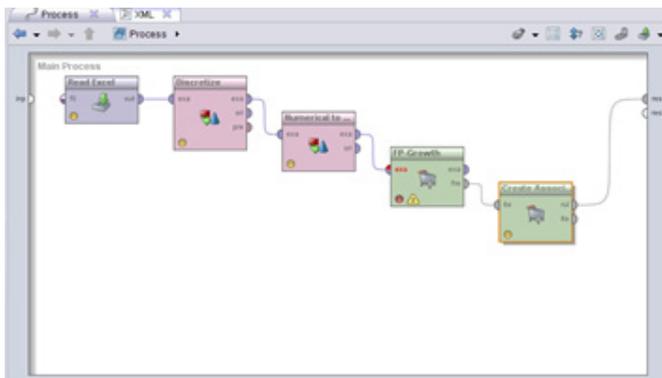

Figure 1: Data Entry Process Through Rapid Miner Software

Figure 1 is an association data mining model with the Apriori algorithm on the RapidMiner software application. In this study, the analysis used the help of the software to process the results.

CONCLUSION

The finding is that the value of support and confidence is directly proportional to the strength of the link between the elements. The analysis of buying data generates multiple association rules in customer buying patterns on supermarkets. Instead of implementing cross selling techniques and promotional strategies by bundling objects, the consequent association rule can be used by supermarkets for developing promotional strategies based on items that frequently are purchased to boost sales by acquiring shopped items. Balance on items often purchased and arrange layout of items at supermarkets simultaneously. A problem with calculating the purchase transaction data using the prior algorithme is that there is an optimization of the time algorithm in which the aforementioned algorithm requires a thorough database

analysis, particularly for large data analyses, to obtain each item combination. Several methods can be used to increase a priori such as hash, transaction reduction, partitioning, sampling and dynamic itemset counting.

REFERENCES

1. Djamaludin, I.; Nursikuwagus, A. (2017). Analisis Pola Pembelian Konsumen Pada Transaksi Penjualan Menggunakan Algoritma Apriori, Simetris : Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer, Vol. 8, No. 2, 671. doi:10.24176/simet.v8i2.1566

2. [2] Jayapana, R. D.; Rahayu, Y. (2015). Analisis pola pembelian konsumen dengan algoritma apriori pada apotek rahayu jepara, UG Jurnal, Vol. Vol. 8, 1–6

3. Hartama, D.; Perdana Windarto, A.; Wanto, A. (2019). The Application of Data Mining in Determining Patterns of Interest of High School Graduates, Journal of Physics: Conference Series, Vol. 1339, No. 1. doi:10.1088/1742-6596/1339/1/012042

4. Windarto, A. P.; Indriani, U.; Raharjo, M. R.; Dewi, L. S. (2020). Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia), Jurnal Media Informatika Budidarma, Vol. 4, No. 3, 855. doi:10.30865/mib.v4i3.2312

5. Jatnika, H.; Huda, M.; Amelia, R. R.; Manuhutu, M. A. (2021). Analysis of Data Mining in the Group of Water Pollution Areas using the K-Means Method in Indonesia Analysis of Data Mining in the Group of Water Pollution Areas using the K-Means Method in Indonesia, Journal of Physics: Conference Series PAPER. doi:10.1088/1742-6596/1783/1/012014

6. Aprianti, W.; Hafizd, K. A.; Rizani, M. R. (2017). Implementasi Association Rules dengan Algoritma Apriori pada Dataset Kemiskinan, Limits: Journal of Mathematics and Its Applications, Vol. 14, No. 2, 57. doi:10.12962/limits.v14i2.2933

Lukman Samboteng, et al. - Market basket analysis of administrative patterns data of consumer purchases using data mining technology

iipp

7. Author, F.; Author, S.; Author, T. (2021). Analysis of Library Book Borrowing Patterns Using Apriori Association Data Mining Techniques (Case : SMK ABC). doi:10.18421/TEMxx-xx

8. Ge, Z.; Song, Z.; Ding, S. X.; Huang, B. (2017). Data Mining and Analytics in the Process Industry: The Role of Machine Learning, IEEE Access, Vol. 5, 20590–20616. doi:10.1109/ACCESS.2017.2756872

9. Aprilla Dennis. (2013). Belajar Data Mining dengan RapidMiner, Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2, Vol. 5, No. 4, 1–5

10. Wahidin, W.; Mugihartadi, M.; Hasanah, T.; Aviani, B.; Pratiwi, H. (2021). Application of data mining techniques using the K-Means Method on Unmet Need of Health Services by Province in Indonesia Application of data mining techniques using the K-Means Method on Unmet Need of Health Services by Province in Indonesia, Journal of Physics: Conference Series PAPER. doi:10.1088/1742-6596/1783/1/012012

11. Elsi, Z. R. S.; Pratiwi, H.; Efendi, Y.; Rusdina, R.; Alfah, R.; Windarto, A. P.; Wiza, F. (2020). Utilization of Data Mining Techniques in National Food Security during the Covid-19 Pandemic in Indonesia, Journal of Physics: Conference Series, Vol. 1594, No. 1. doi:10.1088/1742-6596/1594/1/012007

12. Anil Kumar, K. M.; Ajay, B.; Shashank, R.; Amogha Subramanya, D. A. (2019). An apriori method for topic extraction from text files, International Journal of Recent Technology and Engineering, Vol. 8, No. 2, 2516–2521. doi:10.35940/ijrte.A3068.078219

13. Bailey-Wilson, J. E.; Brennan, J. S.; Bull, S. B.; Culverhouse, R.; Kim, Y.; Jiang, Y.; Jung, J.; Li, Q.; Lamina, C.; Liu, Y.; Mägi, R.; Niu, Y. S.; Simpson, C. L.; Wang, L.; Yilmaz, Y. E.; Zhang, H.; Zhang, Z. (2011). Regression and data mining methods for analyses of multiple rare variants in the Genetic Analysis Workshop 17 mini-exome data, Genetic Epidemiology, Vol. 35, No. SUPPL. 1, 92–100. doi:10.1002/gepi.20657

14. Agustiani, N.; Suhendro, D.; Apriori, K.; Pembahasan, A. (2020). Penerapan Data Mining Metode Apriori Dalam Implementasi Penjualan Di Alfamart, Vol. 2, 300–304

15. Waluyo, A.; Jatnika, H.; Permatasari, M. R. S.; Tuslaela, T.; Purnamasari, I.; Windarto, A. P. (2020). Data Mining Optimization uses C4.5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses, IOP Conference Series: Materials Science and Engineering, Vol. 874, No. 1, 1–9. doi:10.1088/1757-899X/874/1/012024

16. Supriyadi, B.; Windarto, A. P.; Soemartono, T.; Mungad. (2018). Classification of natural disaster prone areas in Indonesia using K-means, International Journal of Grid and Distributed Computing, Vol. 11, No. 8, 87–98

17. Widyastuti, M.; Fepdiani Simanjuntak, A. G.; Hartama, D.; Windarto, A. P.; Wanto, A. (2019). Classification Model C.45 on Determining the Quality of Custumer Service in Bank BTN Pematangsiantar Branch, Journal of Physics: Conference Series, Vol. 1255, No. 012002, 1–6. doi:10.1088/1742-6596/1255/1/012002

18. Jantan, H.; Jamil, A. Z. M. (2019). Association rule mining based crime analysis using apriori algorithm, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, No. 1.5 Special Issue, 18–24. doi:10.30534/ijatcse/2019/0581.52019

19. Yulanda, R. D.; Wahyuningsih, S.; Amijaya, F. D. T. (2019). Association rules with apriori algorithm and hash-based algorithm, Journal of Physics: Conference Series, Vol. 1277, No. 1. doi:10.1088/1742-6596/1277/1/012048

20. Kristiana, T.; Putri, S. A.; Nurmalasari; Handayani, R. I.; Merlina, N.; Yunita, N. (2020). Association Rule Implementation Using Algorithm Apriori to Analize Fishing Pattern in Indonesia, Journal of Physics: Conference Series, Vol. 1641, No. 1. doi:10.1088/1742-6596/1641/1/012072

21. Ruswati, R.; Gufroni, A. I.; Rianto, R. (2018). Associative Analysis Data Mining Pattern Against Traffic Accidents Using Apriori Algorithm, Scientific Journal of Informatics, Vol. 5, No. 2, 91–104. doi:10.15294/sji.v5i2.16199

22. Hanguang, L.; Yu, N. (2012). Intrusion Detection Technology Research Based on Apriori Algorithm, Physics Procedia, Vol. 24, 1615–1620. doi:10.1016/j.phpro.2012.02.238