

Implementation of Apriori Algorithm Data Mining for Increase Sales

Reza Alfianzah¹⁾, Rani Irma Handayani²⁾, Murniyati³⁾

¹⁾²⁾STMIK Nusa Mandiri, Indonesia³⁾Bina Sarana Informatika University, Indonesia

¹⁾rezaalfianzah88@gmail.com, ²⁾ rani.rih@nusamandiri.ac.id, ³⁾ murni.mni@bsi.ac.id

Submitted : Aug 21, 2020 | **Accepted** : Oct 3, 2020 | **Published** : Oct 23, 2020

Abstract— Any company or organization that wants to survive needs to determine the right business strategy. The product sales data carried out by Lakoe Dessert Pondok Kacang will eventually result in a pile of data, so it is unfortunate if it is not re-analyzed. The products offered vary with a wide variety of products as many as 45 products, to find out the products with the most sales and the relationship between one product and another, one of the algorithms is needed in the data mining algorithm, namely the a priori algorithm to find out, and with the help of the Rapidminer 5 application, with a support value 2,4% and a confidence value 50%, products that customers often buy or are interested in can be found. This study used sales data for March 2020, which amounted to 209 transaction data. From the research, it was found that the item with the name Pudding Strawberry and Pudding Vanilla was the product most purchased by consumers. With knowledge of the most sold products and the patterns of purchasing goods by consumers, Lakoe Dessert Pondok Kacang can develop marketing strategies to market other products by analyzing the profits from selling the most sold products and anticipating running out or empty of stock or materials at a later date.

Keywords—Data Mining, Apriori Algorithm, Sales Data, Rapidminer, Association Rule

INTRODUCTION

In the business world, every company must be able to compete and think about how their company can continue to grow and expand its business scale. In order to increase sales of the products they sell, business actors must have various strategies that must be carried out. One way is to use all sales transaction data owned by the company itself. The use of data contained in the information system to support decision-making activities, it is not enough to rely on operational data alone, but also requires a data analysis to extract important information contained therein. The decision makers are trying to take advantage of the data warehouse they already have to get useful information to help make decisions (Oktaviani et al., 2019).

The use of data contained in the information system to support decision-making activities, it is not enough to rely on operational data alone, but also requires a data analysis to extract important information contained therein. The decision makers are trying to take advantage of the data warehouse they already have to get useful information to help make decisions (Badrul, 2016).

With sales activities every day, the existing sales data will increase. The sales data not only serves as archives and Company reports, but the sales data can be used and processed into useful information for the company's marketing strategy (Firmansyah & Merlina, 2020). However, at Lakoe Dessert Pondok Kacang it does not utilize existing sales transaction data for bubble drink, silky drink and silky pudding and the sales transaction data is only used as an archive or report, resulting in data accumulation of unknown benefits.

Sometimes the results of processing data in a simple way do not get the desired results because of the large amount of data being processed. Therefore we need a new data processing method that can provide strategic and accurate information, namely data mining (Rahmawati & Merlina, 2018).

The priori algorithm is a data mining algorithm that provides information on the relationship between data items in the database. The priori algorithm can be used in the sales process. A priori algorithm data mining can help in the sales process by providing a relationship between sales transaction data that has been carried out by customers in order to obtain customer buying patterns and the customers often buy or are interested in. Businesses can use this information to take appropriate business action (Badrul, 2016).

*name of corresponding author



Based on the above background, this study discusses data mining by classifying bubble drink, silky drink and silky pudding brands that are often purchased by customers using apriori algorithm association rules to determine the most sold product, pattern of item and itemset combination of bubble drink, silky drink and silky pudding sales data transactions at Lakoe Dessert Pondok Kacang for one month.

LITERATURE REVIEW

Understanding data mining

According to Retno Tri Vlandari data mining is a series of processes to explore added value in the form of information that has not been known manually from a database. The information generated is obtained by extracting and extracting important or interesting patterns from the data contained in the database. Data mining is mainly used to find knowledge contained in large databases, so it is often called knowledge discovery databases (KDD) (Vlandari, 2017).

According to Sigit Adinugroho and Yiuta Arum Sari, data mining discusses extracting or collecting useful information from data sets. Information that is usually collected is hidden patterns in data, relationships between data elements, or model creation for data forecasting purposes (Adinugroho & Sari, 2018).

Several definitions are also explained according to Lailil Muflikhah, et al that (Muflikhah et al., 2018):

- Data mining is the decomposition (which is not simple) of a set of data into information that has implicit (unreal / clear) potential that was previously unknown.
- Data mining is extracting and analyzing, using automatic or semi-automatic tools, from large amounts of data aimed at finding meaningful patterns.
- Data mining is also part of knowledge discovery databases (KDD).

Data Mining Stages Process

According to tan data mining is one of a series of knowledge discovery databases (KDD). KDD deals with integration techniques and scientific discovery, interpretation and visualization of patterns of data. This series of processes has the following stages (Vlandari, 2017) :

- Data cleaning is the process of removing inconsistent data and noise.
- Data integration is the process of combining data from several sources.
- Data transformation is the process of data being converted into a form suitable for screening.
- The application of data mining techniques is the process of extracting patterns from existing data.
- Evaluation of found patterns is a process of interpreting patterns into knowledge that can be used to support decision making.
- Knowledge presentation is the process of presenting knowledge in a form that is easy for users to understand.

Definition of Algorithm Apriori

According to retno tri vlandari a priori algorithm, including the type of association rules in data mining. Apart from a priori, which is included in this group are generalized rule induction methods and hash-based algorithms. A rule that states the association between attributes is often called an affinity analysis or market basket analysis. Association analysis or association rule mining is a data mining technique to find association rules between a combination of items (Vlandari, 2017).

The importance of an association rule can be determined by two parameters, support (support value), namely the percentage of the combination of these items in the database and confidence (certainty value), which is the strength of the relationship between items in an association rule (Vlandari, 2017).

Definition of Association Rule

Association rule Is a bond or bias, also called a "what to what" relationship. Association rules are also often called market basket analysis (shopping basket analysis), or can be called association rules in the form of "if-then" or "if-then". This rule is calculated from a set of data that is probability (J. L. Putra et al., 2019).

According to retno tri vlandari a priori stated that The basic methodology of association analysis consists of two, namely: analysis of high frequency patterns and the Establishment of associative rules (Vlandari, 2017).

- analysis of high frequency patterns

This stage looks for a combination of items that meet the minimum requirements of the support value. The support value of an item is obtained by the following Equation (1).

$$Support(A) = \frac{\text{Total Transaction Contain A}}{\text{Total Transaction}} * 100\% \quad (1)$$

While the support value of 2 items is obtained from the following Equation (2).

$$Support(A \cap B) = \frac{\text{Total Transaction Contain A dan B}}{\text{Total Transaction}} * 100\% \quad (2)$$

b) Establishment of associative rules

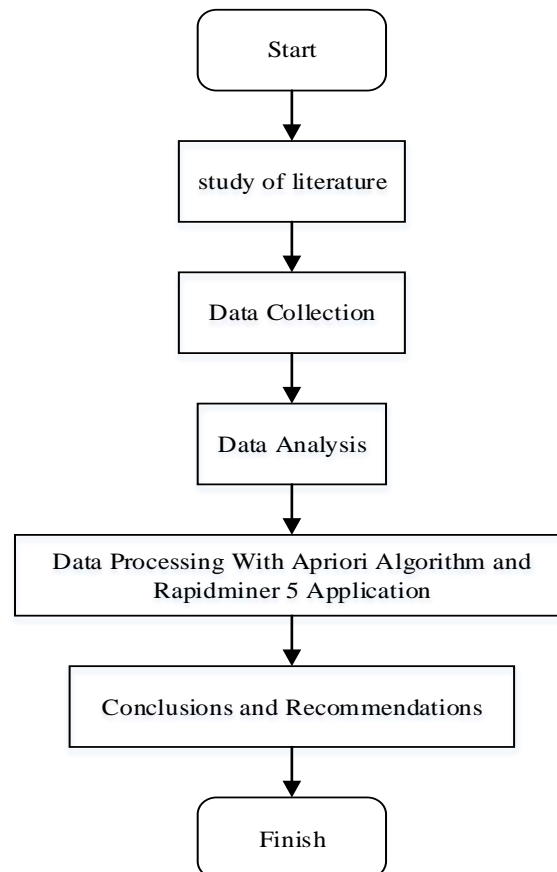
After all high frequency patterns are found, then look for an associative rule that meets the minimum requirements for confidence by calculating the associative confidence rule $A \rightarrow B$ obtained from the following Equation (3).

$$Confidence (B|A) = \frac{\text{Total Transaction Contain A dan B}}{\text{Total Transaction Contain A}} * 100\% \quad (3)$$

PROPOSED METHOD

Research Stages

In this research stage, a systematic research stage is used to help the research be well directed (Firmansyah & Merlina, 2020).



Source : Alfianzah (2020)

Fig 1. Research Stages

Data Collection Methods, Population and Research Samples

As for the method of data collection, collection and method of research that the author will do.

a) Data Collection Methods

Data collection was carried out to obtain the information needed to achieve the research objectives. Data collection will be carried out in the following steps:

- a. Primary data, the authors get data obtained directly from the field, namely through interviews and observations.
- 1) Interview, the author collects data and information through interviews and direct discussions with the Lakoe Dessert Pondok Kacang.

- 2) Observation, the author collects data by recording directly on the object of research to obtain the necessary data in the study. To find out the information needed, researchers collected sales report data in March 2020 at Lakoe Dessert Pondok Kacang
- b. Secondary Data, the authors collect data and information through secondary literature studies, namely data obtained through reference books, books, journals, e-books and other information related to the problem under study.
- b) Population
Population is a generalization area consisting of objects or areas that have certain qualities and characteristics that are determined by researchers to study and then draw conclusions (Taufiqurrahman et al, 2017).
Based on the above conclusions, the population can be defined as the whole object that is the focus of the research. The population in this study is a sales transaction data report that occurred in March 2020.
- c) Research Samples
The sample is part of the number and characteristics of the population. The sample is carried out if the population is large and the researcher is not possible to study everything in the population (Taufiqurrahman et al., 2017). The sample in this research is sales transaction data for bubble drink, silky drink and silky pudding that is successful as many as 209 transaction data.

Data Analysis Method

In this study, the author uses quantitative data analysis, which is a data analysis that can be used if the conclusions obtained can be proven by numbers and a formula that is related to the research analysis is used in calculations. In this case a priori analysis is used.

The writer uses a priori algorithm calculation by testing the hypothesis of two variables between the value of support and the value of confidence. After that, a priori algorithm is calculated using Rapidminer 5 which functions as a matching of the results obtained in the previous calculation.

Data processing uses a priori algorithm calculations.

a) High Frequency Pattern Analysis

This stage looks for a combination of items that meet the minimum requirements of the support value is 2,4%. The support value for 1 itemset can be obtained using the Equation (1). While the support value of 2 items can be obtained using the Equation (2). The search for high frequency patterns will be stopped if the combination does not meet the specified Support value.

b) Establishment of associative rule

After all high frequency patterns are found, then look for an associative rule that meets the minimum requirements for confidence is 50% by calculating the associative confidence rule $A \rightarrow B$ can obtained using the Equation (3).

c) Establishment of final association rules

The final association rules are ordered based on the minimum support and minimum confidence that has been determined

RESULT AND DISCUSSIONS

The following is a list of bubble drink products, silky drinks and silky pudding in the pondok kacang lakoe dessert can be seen in the Table 1. List of Products Bubble Drink, Silky Drink and Silky Pudding.

Table 1.
List of Products Bubble Drink, Silky Drink and Silky Pudding

Number	Product Name
1	Bubblegum
2	Vanilla
3	Red Velvet
4	Pudding Strawberry
5	Choco Hazelnut
6	Choco Oreo
7	Lakoe Choco
8	Taro

9	Vanilla Oreo
10	Coffee Caramel
11	Lakoe Green Tea
12	Strawberry
13	Tiramisu
14	Cappuccino
15	Vanilla Latte
16	Mango Yakult
17	Strawberry Yakult
18	Durian
19	Thai Tea
20	Mango
21	Strawberry Silky Drink
22	Lychee Yakult
23	Avocado
24	Creamy Choco
25	Pudding Mango
26	Pudding Vanilla
27	Pudding Green Tea
28	Pudding Taro
29	Choco Milktea
30	Lakoe Choco Silky Drink
31	Mochaccino
32	Bubblegum Silky Drink
33	Choco Taro
34	Pudding Chocolate
35	Choco Silverqueen
36	Choco Nutela
37	Lakoe Milktea
38	Choco Vanilla
39	Pudding Bubblegum
40	Taro Milktea
41	Hazelnut Milktea
42	Choco Hazelnut Milktea
43	Lakoe Hazelnut
44	Lychee
45	Cappuccino Oreo

Source : Lakoe Deseert Pondok Kacang (2020)

High Frequency Pattern Analysis

1. Forming of 1 itemset

The process of forming C1 or what is called 1 itemset is calculated with a minimum support of 2,4%. so the combination of 1 itemset that does not meet the minimum support will be removed. The support value for 1 itemset can be obtained using the Equation (1). The following is a list of 1 itemset that meets minimum support:

Table 2.
Table 3. Support 1 Itemset

Itemset	Support	Support 1 Itemset
Bubblegum	14/209	0,067 6,70%
Vanilla	11/209	0,053 5,30%
Pudding Strawberry	9/209	0,043 4,30%
Choco Hazelnut	19/209	0,091 9,10%
Choco Oreo	13/209	0,062 6,20%
Lakoe Choco	34/209	0,163 16,30%
Taro	12/209	0,057 5,70%
Vanilla Oreo	11/209	0,053 5,30%
Coffee Caramel	5/209	0,024 2,40%

Lakoe Green Tea	10/209	0,048	4,80%
Strawberry	5/209	0,024	2,40%
Tiramisu	5/209	0,024	2,40%
Cappuccino	16/209	0,077	7,70%
Vanilla Latte	12/209	0,057	5,70%
Mango Yakult	10/209	0,048	4,80%
Strawberry Yakult	5/209	0,024	2,40%
Durian	5/209	0,024	2,40%
Thai Tea	11/209	0,053	5,30%
Mango	15/209	0,072	7,20%
Avocado	6/209	0,029	2,90%
Creamy Choco	11/209	0,053	5,30%
Pudding Mango	6/209	0,029	2,90%
Pudding Vanilla	7/209	0,033	3,30%
Pudding Green Tea	5/209	0,024	2,40%
Lakoe Choco Silky Drink	7/209	0,033	3,30%
Pudding Chocolate	5/209	0,024	2,40%
Choco Silverqueen	6/209	0,029	2,90%
Lakoe Milktea	6/209	0,029	2,90%

Source : Research Result (2020)

2. Forming of 2 itemset

The process of forming C2 or what is called 2 itemset is calculated with a minimum support of 2,4%. The support value for 2 itemset can be obtained using the Equation (2).

The minimum support specified is 2,4%. After forming 2 itemsets, 378 combinations of 2 itemsets were formed. a combination of 2 itemset that does not meet the minimum support will be removed. can be seen in the table 3. Support 2 itemset.

Table 3
Support 2 Itemset

Itemset	Support	Support 2 Itemset
Pudding Strawberry, Pudding Vanilla	5/209	0,024 2,40%

Source : Research Result (2020)

Because you can only get 1 combination of 2 itemset then the calculation will not continue to C3 or 3 itemset and stop on the calculation of the combination of 2 itemset or C2.

3. Establishment of Association Rules

After all high frequency patterns are found, then look for association rules that meet the minimum requirements for confidence by calculating the confidence or association $A \rightarrow B$, with a minimum confidence of 50%. The confidence value can be obtained using the Equation (3).

So the formation of association rules or confidence that does not meet the minimum confidence will be eliminated. Calculation of association rules or confidence between products can be seen in Table 4. Confidence.

Table 4.
Table 5. Confidence

Rules	Calculation	Confidence
if you buy Pudding Strawberry then you will buy Pudding Vanilla	5/9	0,556 55,60%
if you buy Pudding Vanilla then you will buy Pudding Strawberry	5/7	0,714 71,40%

Source : Research Result (2020)

4. Association rules Final

The final association rules are ordered based on the minimum support and minimum confidence that has been determined, can be seen in the table 5. Association rules final.

Table 5
Association Rules Final

Rules	Support	Confidence
if you buy Pudding Strawberry then you will buy Pudding Vanilla	0,024 2,40%	0,556 55,60%
if you buy Pudding Vanilla then you will buy Pudding Strawberry	0,024 2,40%	0,714 71,40%

Source : Research Result (2020)

Based on the table of the final association rules above, it is known that bubble drink, silky drink and silky pudding products, the most sold are strawberry pudding and vanilla pudding, with the knowledge that the most sold products are, so that companies can develop marketing or promotional strategies to market products with other products. by examining what the advantages of the most-sold product are with other products and anticipating running out or emptying stock or materials in the future.

5. Implementation of the apriori algorithm on Rapidminer 5

The a priori algorithm implementation in this study uses the Rapidminer 5 application for testing.

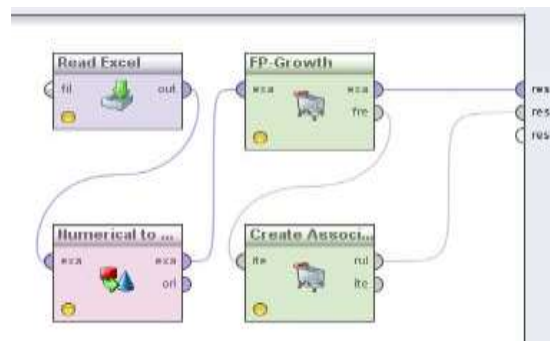


Fig 2. Main Process

In Figure 2. The Main Process is formed through the process steps by connecting each operator, namely: Read Excel connecting to Numerical to Binomial, Numerical to Binomial connecting to Fp-Growth, Fp- Growth connecting to Result Process, Fp- Growth connecting to create Association Rules and Create Association Rules connecting to the Results Process.

first on the read excel operator, click on the import configuration wizard parameter to import transaction data that is tabular. after that in the FP-Growth operator, in the parameter, the writer must fill in the minimum support of 0.024. then in the Create Association Rules operator, in the parameters, the writer will fill in minimum confidence 0.5. finally click run and the calculation results will appear.

Then the results of the Rule are shaped into 2 rules from the results of Rapidminer 5 as follows:

No.	Premises	Conclusion	Support	Confidence
1	Pudding Strawberry	Pudding Vanilla	0.024	0.556
2	Pudding Vanilla	Pudding Strawberry	0.024	0.714

Fig 3. Table View

In Figure 3. Table View from the overall results of the a priori algorithm implementation process in manual calculations and calculations in Rapidminer 5 for bubble drink, silky drink and silky pudding sales transaction data as many as 209 data by providing a minimum limit of 0.024 Support and 0.5 Confidence that has been set, so as to get the results of rules 2 as follows:

1. Pudding Strawberry, Pudding Vanilla with a Support Value of 0.024 and a Confidence Value of 0.556.
2. Pudding Vannila, Pudding Strawberry with a Support Value of 0.024 and a Confidence Value of 0.714.

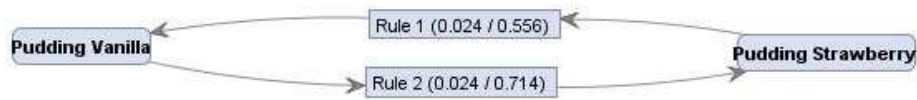


Fig 4. Graph View

In Figure 4. It can be concluded from the value of the most superior rules with 0,024 Support and 0,714 Confidence is if you buy Pudding Vannila then you will buy Pudding Strawberry.

CONCLUSION

Based on the research conducted by the author, the author can draw the conclusion that the sales of Bubble Drink, silky drink and Silky Pudding products are the most sold at Lakoe Dessert Pondok Kacang can be found using the apriori algorithm, by looking at products that meet minimum support of 2,4% and minimum confidence of 50%. The most sold products are Strawberry Pudding and Vanilla Pudding with the conclusion that, Strawberry Pudding and Vanilla Pudding products with a support value of 0.024 or 2,4% and a confidence value of 0.556 or 55,6%. or Vanilla Pudding and Strawberry Pudding products with a support value of 0.024 or 2,4% and a confidence value of 0.714 or 71,4%. Then after testing by manual calculation and using Rapidminer 5 software, it can be seen that the results obtained from manual calculations are the same as calculations using Rapidminer 5 software, so in general the software has worked well because the calculation process is as expected.

SUGGESTION

Based on the research conducted by the author, the suggestion are as follows :

- That by knowing the products that are often sold, the management of Lakoe Dessert Pondok Kacang can predict future market needs, and can calculate what stock of goods that must be multiplied because there are many enthusiasts, and what items must be reduced because of lack of interest, So that by knowing the pattern of purchasing these goods, it can further increase the turnover of the company.
- The software used to process data in forming association rules for final support and confidence can use other supporting software, not only relying on Rapidminer 5.
- The addition of research objects does not only cover one month but can be up to several months backward so that data is more accurate.
- In using the Apriori Algorithm, an understanding of Association rules and Data Mining and how to run Rapidminer software applications is needed.
- The use of the Apriori Algorithm is very practical but still needs to be compared with other algorithms, to test the extent to which the Apriori Algorithm can still be relied on to process and find patterns of relationships (associations) between items in a large database

REFERENCES

- Adib, H. S. (2017). Teknik Pengembangan Instrumen Penelitian Ilmiah Di Perguruan Tinggi Keagamaan Islam. *Seminar Nasional Pendidikan, Sains Dan Teknologi*, 139–157. Retrieved from <https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/3054>
- Adinugroho, S., & Sari, Y. A. (2018). *Implementasi Data Mining Menggunakan WEKA (Pertama)*. Malang: UB Press.
- Adiwihardja, C., Hardi, N., & Widyastuty, W. (2019). Implementasi Data Mining Penjualan Kosmetik Pada Toko Zahrani Menggunakan Algoritma Apriori, *11(2)*, 1–7. Retrieved from <http://speed.web.id/ejournal/index.php/speed/article/view/427>
- Andriyani, D., Harahap, E., Badruzzaman, F. H., Fajar, M. Y., & Darmawan, D. (2019). Aplikasi Microsoft Excel Dalam Penyelesaian Masalah Rata-rata Data Berkelompok Microsoft Excel Application in Solving The Average of Group Data Problems, *18(1)*, 41–46. Retrieved from <https://ejournal.unisba.ac.id/index.php/matematika/article/view/5078>
- Badrul, M. (2016). Algoritma asosiasi dengan algoritma apriori untuk analisa data penjualan. *Pilar Nusa Mandiri, XIII(2)*, 121–129. Retrieved from <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/266/236/>
- Cep, A., Murni, C., & Reynida, H. (2018). Implementasi Data Mining Penjualan Tas Pada Toko Fabella Shop Menggunakan Algoritma Apriori, *10(4)*, 92–97. Retrieved from <http://speed.web.id/ejournal/index.php/speed/article/view/416>
- Firmansyah, A., & Merlina, N. (2020). Prediksi Pola Penjualan Tiket Kapal Pt. Pelni Cabang Makassar Menggunakan Metode Algoritma Apriori. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *5(2)*,

- 183–190. <https://doi.org/10.33480/jitk.v5i2.1123>
- Muflikhah, L., Ratnawati, D. E., & Putri, R. R. M. (2018). *Data Mining*. Malang: UB Press.
- Oktaviani, A., TM Napitupul, G., Sarkawi, D., & Yulianti, I. (2019). Penerapan Data Mining Terhadap Penjualan Pipa Pada Cv. Gaskindo Sentosa Menggunakan Metode Algoritma Apriori. *Jurnal Riset Informatika*, 1(4), 167–172. <https://doi.org/10.34288/jri.v1i4.96>
- Putra, A. A. C., Haryanto, H., & Dolphina, E. (2018). Implementasi Metode Association Rule Mining Dengan Algoritma Apriori Untuk Rekomendasi Promo Barang. *Csrid*, 11, 89–99. Retrieved from <http://csrid.potensi-utama.ac.id/index.php/CSRID/article/view/277>
- Putra, J. L., Raharjo, M., Sandi, T. A. A., Ridwan, R., & Prasetyo, R. (2019). Implementasi Algoritma Apriori Terhadap Data Penjualan Pada Perusahaan Retail. *Jurnal Pilar Nusa Mandiri*, 15(1), 85–90. <https://doi.org/10.33480/pilar.v15i1.113>
- Rahmawati, F., & Merlina, N. (2018). Metode Data Mining Terhadap Data Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, 6(1), 9–20. <https://doi.org/10.33558/piksel.v6i1.1390>
- Rulianto, J., & Mustika, W. P. (2019). Implementasi Algoritma Apriori Terhadap Data Penjualan Oli Mobil Berbasis Sistem Pencarian Aturan Asosiasi. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 3(2), 316. <https://doi.org/10.30645/j-sakti.v3i2.150>
- Sintia, S., Poningsih, P., Saragih, I. S., Wanto, A., & Damanik, I. S. (2019). Penerapan Algoritma Apriori Dalam Memprediksi Hasil Penjualan Sparepart PC (Studi Kasus : Toko Sentra Computer). *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 910. <https://doi.org/10.30645/senaris.v1i0.99>
- Tamba, S. J., & Bu, E. (2019). Implementasi Algoritma Apriori Pada Sistem Persediaan Buah-Buahan (Studi Kasus : Lotte Mart Wholesale Medan), 18, 616–621. Retrieved from <https://www.ejurnal.stmik-budidarma.ac.id/index.php/pelita/article/view/1842>
- Taufiqurrahman, F., Retnowati, N., & Negoro, B. K. (2017). Faktor Yang Mempengaruhi Keputusan Pembelian Tidak Terencana Pada Minimarket Greensmart Di Kota Baru Driyorejo-Gresik. *Journal of Manajemen*, 3(3). <https://doi.org/10.1017/CBO9781107415324.004>
- Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer* (Cetakan 1). Yogyakarta: Gava Media.