International Journal Software Engineering and Computer Science (IJSECS) 3 (3), **2023, 449-456**
Published Online December 2023 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v3i3.1877.

Analysis of the Apriori Algorithm for Enhancing Retail Product Staple Sales Recommendations

Avip Kurniawan
Informatics Engineering Study Program, Faculty of Engineering, Universitas Krisnadwipayana, Bekasi Regency, West Java Province, Indonesia.
Email: avip.kurniawan@unkris.ac.id

Niko Suwaryo *
Digital Business Study Program, Faculty of Social Sciences and Technology, Universitas Medika Suherman, Bekasi Regency, West Java Province, Indonesia.
E-mail: niko@medikasuherman.ac.id

Received: 30 November 2023; Accepted: 10 December 2023; Published: 20 December 2023.

Abstract: Products are fundamental commodities in the market that cater to various consumer needs and desires. This research employs the Apriori algorithm to generate product recommendations based on the analysis of high-demand patterns arising from product sales and association patterns. Specifically, we focus on identifying elevated sales in categories such as Bulk Products, Biscuits/Snacks, Drinks, Milk/Coffee/Tea, and Sauces & Spices during specific time intervals. The model's evaluation and validation entail measuring the Lift Ratio value, a key metric. In our assessment using the RapidMiner Studio application, we find that the Lift Ratio value equals 1. Consequently, our model asserts that combinations with a Lift Ratio value greater than or equal to 1 are deemed valid and beneficial.

Keywords: Retail Products; Data Mining; Apriori Algorithm; Product Recommendations; Fundamental Ingredients.

1. Introduction

In an increasingly digitally connected era, the use of computer technology has become an integral part of various aspects of our lives. Past research has highlighted the broad and varied applications of computer technology, providing valuable insight into how this technology can impact and improve various fields, including business and commerce. One of the relevant previous studies is research by Wali and Ahmad [1]. They explore the use of computer technology in learning software development through designing source code library applications. This research shows the great potential of computer technology in supporting education and skills development in the field of software development. On the other hand, research by Munawir and Ardiansyah (2017) highlights the use of computer technology in the context of human resource management. They developed a Decision Support System for the selection of extraordinary employees at the Bank Indonesia Representative Office, Aceh [2]. This research shows that computer technology can provide valuable support in critical decision-making processes. In addition, research by Satria, Yana, Munadi, and Syahreza (2017) presents the application of computer technology in environmental monitoring with the development of a web-based Real-Time Flood Early Warning System. This highlights the role of computer technology in natural disaster risk mitigation and environmental protection [3]. The use of computer technology has also become the basis for various innovations, such as fast computer startups [4], paradigm changes in philosophy [5], the use of computer games in education [6], and the development of computer networks [7]. This research reflects how computer technology continues to change the landscape of various disciplines. In this research, we will explore the use of computer technology in analyzing retail products and providing relevant recommendations. By referring to this previous research, we seek to understand how this technology can make a positive contribution in the context of business and retail trade. As such, this research will build on the strong foundations that have been developed by previous researchers, with the hope of providing valuable insights for strategy and decision development in this area.

The competition that occurs in the business world forces business people to always think about strategies and breakthroughs that can ensure the continuity of the business they run. Products are basic materials that can be offered to a market and can satisfy a want or need [1]. At the retail level, products are often called merchandise. In manufacturing, products are purchased in raw form and sold as finished goods. Products in the form of raw goods such as metal or agricultural products are often called commodities. Products include all activities that involve selling goods directly to

end consumers for personal, non-business use. Searching for sales patterns begins with processing item data from each purchase, then looking for relationships between the items purchased. Searching for this information is almost the same as looking for opportunities for the appearance of purchased goods according to people's shopping habits and the number of existing transactions. The a priori algorithm functions to form candidate combinations of item sets.

The problem that often occurs is a decrease in sales, uncertainty about the goods being sold means that income is not optimal. This is due to the accumulation of unsold products, products that are not sold by consumers. Therefore, an analysis is needed that can predict product needs or recommendations for the coming period. Data processing is a term used to describe the discovery of knowledge in databases. Data mining is the process of discovering meaningful relationships, patterns, and trends by examining large sets of data stored in storage using pattern recognition techniques, one of which is Association rules Mining is an excellent way to provide scientific decision support to the market through association mining. relationships between goods that have been distributed together. The discovery of this relationship can help distributors to develop sales strategies by considering items that customers frequently purchase together. In this research we analyze retail product recommendations by looking at distributor demand so that with knowledge and information related to product recommendations we can anticipate unavailability of goods.

2. Research Method

In conducting this research, a systematic approach was followed to ensure methodical progression and consistency. The following outlines the key steps taken during the study:

2.1. Problem Identification

The initial phase involved identifying the problem and establishing its boundaries. The primary objective was to gain a comprehensive understanding of the problem's scope, define clear goals, and ascertain the potential benefits to be derived from its resolution.

2.2. Problem Analysis

In this step, a rigorous analysis of the identified problem was undertaken to delve deep into its nuances. The goal was to acquire a profound understanding of the problem's intricacies, enabling a well-informed approach to its resolution.

2.3. Data Collection

Data acquisition played a pivotal role in supporting the research. The following methods were employed for data collection:

1) Field Study: This involved conducting interviews with relevant stakeholders and making on-site observations to gain insights into the real-world context. Additionally, pertinent company data related to the research was collected.
2) Literature Review: Extensive study of existing literature, encompassing books, journals, and prior research relevant to the research topic, was conducted. This literature review aimed to establish a strong knowledge base aligned with the research's objectives.

2.4. Data Processing

Data processing in this research adhered to the principles of Knowledge Discovery in Databases (KDD). It encompassed the transformation and preparation of raw data for subsequent analysis.

2.5. Data Analysis

The data mining analysis process was executed utilizing the Apriori algorithm. This process was instrumental in uncovering patterns and relationships among items within the dataset.

2.6. Implementation

Based on the outcomes of the data analysis, the author proceeded to implement the Apriori algorithm using RapidMiner software, operating within the Windows environment. This implementation aimed to translate the research findings into practical applications.

3. Result and Discussion

3.1 Results

3.1.1 Test Results

In this research the author used testing with the RapidMiner Studio application. In this process, association methods and a priori algorithms are applied to find itemset relationships with proper accuracy. Test results are obtained by carrying out the following steps:

1) In the RapidMiner Studio application, import the required data by clicking Import Data , then select the tabular data set that will be used .
2) Click the Design menu , in the process view, add the tabular dataset in the folder to the process view window.
3) Numerical to Binomial function operator (to change binary to binomial numbers) then drag it to the process view.
4) Next, to apply *the frequency item* , select the FP-Growth function operator in the Associations sub menu, then drag it to the process display window.
5) In the operator parameters of the FP-Growth function , determine the minimum *support value* that will be determined. In this test, the minimum *support value* determined is 0.40 or equal to 40%.
6) Create Association Rule function operator then drag it to the process view.
7) We can specify parameters in the Create Association Rule function operator , namely determining the *min.Confidence value* . In this test, the *Confidence value* determined by the author was 0.75 or equal to 75%. For other parameters, use *the defaults* from the RapidMiner Studio application.

Following steps 1 to 7, the author established connectors between the designed functions in the RapidMiner Studio application, forming a flow diagram as shown in Figure 1 below. The RapidMiner application successfully generated results from the implemented Apriori algorithm. In general, the results of establishing Association rules from testing the RapidMiner Studio application can be seen in the following image (Figure 1).

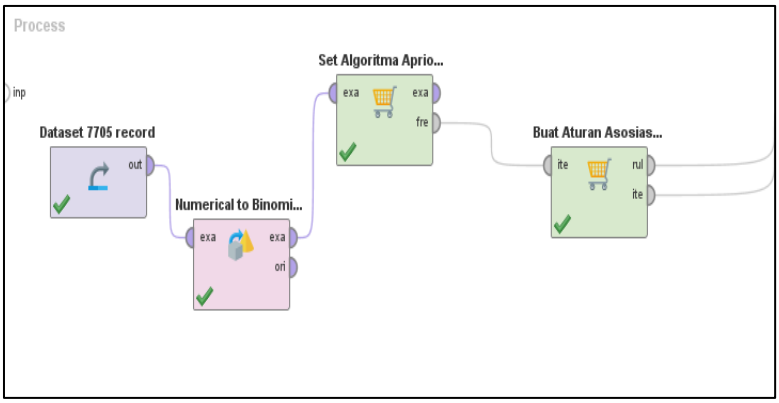

Figure 1. Process model in the RapidMiner Studio application on the dataset used

Next, the RapidMiner application succeeded in producing a result from the implementation of a previously designed model, namely the Apriori algorithm. In general, the results of establishing Association rules from testing the RapidMiner Studio application can be seen in the following image:

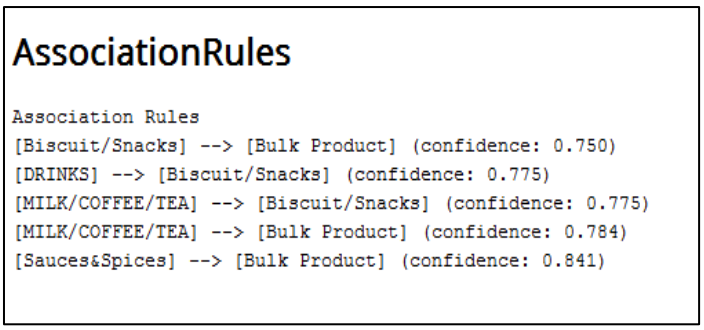

Figure 2. Results of establishing Association rules in the RapidMiner Studio application

3.1.2 Analysis of Results

After experimenting with the application of the Association method and the Apriori algorithm in searching for patterns of association rules for the relationship between material items in sales transaction data, we can produce association rules for each process carried out. The dataset used is 7705 data records and consists of 23 items which will be tested in the process of forming the association method model with the Apriori algorithm. Through testing using the RapidMiner Studio application, the author applies a minimum support value of 0.40 or 40% and a minimum confidence of 0.75 or 75%. support value for each item and of the 23 items only produces 7 items that meet the requirements for the specified minimum support value, namely the Biscuit/Snacks item with a support value of 0.594, then the Bulk Product item which has a support value 0.619, the Drinks item with a support value of 0.540, next the Detergent item with a support value of 0.492, next the H&B item which has a support value of 0.405, the Milk/Coffee/Tea item which has a support value of

0.568 and finally the Sauces&Spices item which has a support value of 0.485. These seven items are a set of k-1 itemsets, and will be candidates for searching for combinations in the next process, namely k-2 itemsets.

Item 1	Support
Biscuit/Snacks	0.594
Bulk Product	0.619
DRINKS	0.540
Detergent	0.492
H&B	0.405
MILK/COFFEE/TEA	0.568
Sauces&Spices	0.485

Figure 3. K-1 itemsets that meet the min value. Support for the RapidMiner Studio application

In calculating the support for k-2 itemset combinations, in this test there are 6 itemset combinations that meet the requirements for achieving the specified minimum *support value*. The following is an image of the results of testing the RapidMiner Studio application.

Item 1	Item 2	Support
Bulk Product	Biscuit/Snacks	0.446
Bulk Product	MILK/COFFEE/TEA	0.445
Bulk Product	DRINKS	0.403
Bulk Product	Sauces&Spices	0.408
Biscuit/Snacks	MILK/COFFEE/TEA	0.440
Biscuit/Snacks	DRINKS	0.419

Figure 4. K-2 itemsets that meet the min value. Support for the RapidMiner Studio application

Next, to see how big the relationship between each item is, in this testing stage the author determined a minimum *confidence value* of 0.75 or 75% in the RapidMiner Studio application. By applying the minimum *confidence value* , the RapidMiner Studio application produces 5 K-2 itemset combination rules that meet the requirements as seen in Figure 5 below.

No.	Premises	Conclusion	Support	Confidence
1	Biscuit/Snacks	Bulk Product	0.446	0.750
2	DRINKS	Biscuit/Snacks	0.419	0.775
3	MILK/COFFEE/TEA	Biscuit/Snacks	0.440	0.775
4	MILK/COFFEE/TEA	Bulk Product	0.445	0.784
5	Sauces&Spices	Bulk Product	0.408	0.841

Figure 5. K-2 itemsets that meet the min value. Confidence in the RapidMiner Studio application

And the following is a graph of the relationship between items that form a combination in the K-2 itemset that meets the min value. *Support* and min value. *Confidence* defined in the RapidMiner Studio application.

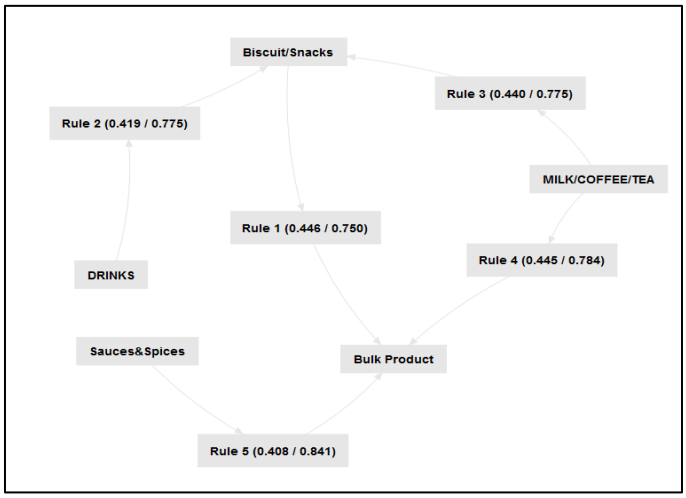

Figure 6. Graph of the relationship between items in the RapidMiner Studio application

Thus, from the existing dataset of 7705 records, after going through the data mining process using the Association method and applying the Apriori algorithm, we can produce a pattern of rules between 5 product category items that are often sold simultaneously in a sales transaction, such as Biscuits/Snacks, **Bulk** Products, **Drinks**, **Milk/Coffee/Tea**, and **Sauces&Spices** . Both one item and another item have a fairly close relationship considering that each combination of these items has a fairly good *confidence value.* The test results using the RapidMiner Studio application also have similarities with the calculation simulations that the author carried out in modeling the Association method and Apriori algorithm on sales transaction data to find a pattern of relationships between product categories. The following is a comparison between the K-1 itemset formation process both carried out in manual simulation and with the RapidMiner Studio application.

Itemset pada Simulasi	Nilai Support	Itemset pada RapidMiner Studio	Nilai Support
Biscuit/Snacks	59,4%	Biscuit/Snacks	0,594
Bulk Product	61,9%	Bulk Product	0,619
Detergent	49,2%	Detergent	0,492
Drinks	54,0%	Drinks	0,540
H&B	40,5%	H&B	0,405
Milk/Coffee/Tea	56,8%	Milk/Coffee/Tea	0,568
Sauces&Spices	48,5%	Sauces&Spices	0,485

Figure 7. Comparison of K-1 itemsets in simulation and application RapidMiner Studio

Due to the similarity of the itemsets produced in the 1st iteration, the formation of the K-2 itemset combination also has similar results as in Figure 7.

3.1.3 Model Evaluation

Model evaluation and validation is carried out by measuring the analysis results to determine the value *L ift R atio* . From all the association rules formed, in the model evaluation test using the RapidMiner Studio application it has a *Lift Ratio value* = 1. Thus, the model applied is if *the Lift Ratio* value *is* > = 1 then the resulting combination is truly valid and has benefits. Details of the lift ratio results can be seen in the following image.

No.	Premises	Conclusion	Lift
1	Biscuit/Snacks	Bulk Product	1.211
2	DRINKS	Biscuit/Snacks	1.303
3	MILK/COFFEE/TEA	Biscuit/Snacks	1.304
4	MILK/COFFEE/TEA	Bulk Product	1.266
5	Sauces&Spices	Bulk Product	1.357

Figure 8. Lift Ratio value in the RapidMiner Studio application.

3.2. Discussion

After running experiments by applying the Association Method and Apriori Algorithm to search for patterns of association rules in goods sales transaction data, we succeeded in producing several significant association rules. The dataset used consists of 7705 data records and includes 23 items tested in the process of forming the association method model with the Apriori algorithm. During testing using the RapidMiner Studio application, the author applied a minimum support value of 0.40 (40%) and a minimum confidence level of 0.75 (75%). From the calculation results, each item has a different support value. Of the 23 items tested, only 7 items met the specified minimum support value requirements. These items include Biscuits/Snacks (support: 0.594), Bulk Products (support: 0.619), Drinks (support: 0.540), Detergent (support: 0.492), H&B (support: 0.405), Milk/Coffee/Tea (support: 0.568), and Sauces & Spices (support: 0.485). These seven items form the k-1 item set and become candidates for combination search in the next process, namely the k-2 item set. At the support calculation stage for the k-2 item set combination, in this test there are 6 item set combinations that meet the specified minimum support value requirements. Test results using the RapidMiner Studio application can be seen in Figure 4 in the report. To assess how strong the relationship between items is, at this testing stage, the author determined a minimum confidence level value of 0.75 (75%) in the RapidMiner Studio application. This results in 5 rule combinations of k-2 item sets that meet the requirements. Thus, from the existing dataset of 7705 data records, after going through a data mining process using the association method and applying the Apriori algorithm, we succeeded in producing a pattern of rules between 5 product categories that are often sold simultaneously in sales transactions. These categories include Biscuits/Snacks, Bulk Products, Drinks, Milk/Coffee/Tea, and Sauces & Spices. Each item shows a close relationship, considering that every combination of these items has a good trust value.

4. Related Work

To identify and understand marketing strategies in the sale of important consumer products, much research has been conducted in this domain. Ariesya *et al*. (2019) proposed research that focuses on the application of the FP-Growth and Apriori algorithms to determine appropriate marketing strategies in selling consumer staple products. They use data mining to identify patterns of consumer demand for basic needs. The results of this research produce decision rules that can help in determining effective marketing strategies [8]. Research by Wijaya (2022) regarding sales transaction analysis in increasing sales promotions using the WEKA tool has been carried out. This research aims to provide valuable information to a supermarket with more than 50 types of food and non-food products. Through sales data analysis, this research produces rules that can be used by supermarkets as a marketing strategy [9]. Firmansyah (2021) presents a case study regarding shopping basket analysis for book sales promotion at Gramedia Matraman Jakarta. They use the FP Growth algorithm to identify book purchasing patterns that can be used for appropriate promotions. The results create a set of rules that Gramedia can use as a marketing strategy [10].

Additionally, Aurelia *et al*. (2022) created a mobile-based application using the Apriori algorithm as a business development strategy during the COVID-19 pandemic. They use shopping cart analysis methods to analyze customer purchasing behavior and patterns from various e-commerce platforms [11]. Srivastava *et al*. (2023) have developed an automated e-grocery shopping system by utilizing the Apriori algorithm to suggest items to users based on their purchase history. This is an example of applying data mining techniques to improve the online shopping experience [12]. Koomsap *et al*. (2023) presented a knowledge-based approach to managing retail stores and associated warehouses. They use knowledge algorithms to help optimize retail store and related warehouse operations [13]. Apart from that, several other studies have also been carried out to analyze sales patterns and develop marketing strategies using data mining algorithms such as Apriori and FP-Growth [14][15][16].

This research has relevance in data analysis and marketing strategy. This is because this research aims to identify effective marketing strategies based on consumer purchasing patterns, which have a direct impact on business success and increasing product sales. What differentiates this research is the use of a combination of the FP-Growth algorithm and the Apriori Algorithm in analyzing sales transaction data. This combination allows research to uncover purchasing patterns that might go undetected using just one algorithm. FP-Growth is used to identify purchasing patterns that frequently appear in the data, while the Apriori Algorithm is used to generate association rules that can be used to develop more sophisticated marketing strategies. Apart from that, this research is also unique in the different study cases. In the references described previously, there are case studies from various sectors, including sales of important consumer products, books, mobile-based applications, and e-grocery shopping. This shows that the data mining approach used in this research can be widely applied in various industries. The importance of this research lies in its ability to uncover relevant purchasing patterns in sales transaction data and generate effective marketing strategies. The combination of the algorithms used, and the variety of study cases make this research a contribution to the development of data-based marketing science and practice.

5. Conclusion

In conclusion, the application of the Association method to analyze sales transaction data through predefined stages has proven to be effective. It successfully generates new association rules from the sales transaction dataset, which can serve as valuable references for optimizing product management in alignment with customer needs. The utilization of the Apriori algorithm allows us to identify sales patterns of high-demand products based on combinations of product sales. The association patterns reveal that certain product categories, such as Bulk Products, Biscuits/Snacks, Drinks, Milk/Coffee/Tea, and Sauces&Spices, tend to have high sales simultaneously. The testing conducted using the Rapid Miner tool validates the results and demonstrates a linear and proportional correlation with manual calculation model simulations. Importantly, all generated rule patterns exhibit a Lift Ratio value greater than 1, indicating the validity and effectiveness of the model.

References

[1] Wali, M. and Ahmad, L., 2017. Perancangan Aplikasi Source code library Sebagai Solusi Pembelajaran Pengembangan Perangkat Lunak. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *1*(1), pp.39-47. DOI: https://doi.org/10.35870/jtik.v1i1.32.

[2] Munawir, M. and Ardiansyah, A., 2017. Decision Support System Pemilihan Karyawan Berprestasi Dengan Pendekatan Analisa Gap Profile matching Di Kantor Perwakilan Bank Indonesia Provinsi Aceh. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *1*(1), pp.7-14. DOI: https://doi.org/10.35870/jtik.v1i1.28.

[3] Satria, D., Yana, S., Munadi, R. and Syahreza, S., 2017. Sistem peringatan dini banjir secara real-time berbasis web menggunakan arduino dan ethernet. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *1*(1), pp.1-6. DOI: https://doi.org/10.35870/jtik.v1i1.27.

[4] Iyigun, M., Bak, Y., Wilson, EN, Stark, KV, Zhang, S., Stemen, PL, King, BE, Karagounis, V. And Jain, N., Microsoft Technology Licensing LLC, 2019. Fast Computer Startups. US Patent 10,268,487.

[5] Sloman, A., 2019. *The computer revolution in philosophy: Philosophy, science and models of mind*. Author.

[6] Mayer, R.E., 2019. Computer games in education. *Annual review of psychology*, *70*, pp.531-549.

[7] Davie, BS And Peterson, LL, 2019. Computer Networks. Morgan Kaufmann.

[8] Ariestya, W.W., Supriyatin, W. and Astuti, I., 2019. Marketing strategy for the determination of staple consumer products using FP-growth and apriori algorithm. *Jurnal Ilmiah Ekonomi Bisnis*, *24*(3), pp.225-235. DOI: http://dx.doi.org/10.35760/eb.2019.v24i3.2229.

[9] Wijaya, H.O.L., 2022. Analysis of Sales Transactions in Increasing Sales Promotion Using the WEKA Tool: Datamining. *Adpebi Science Series*, pp.1-8. DOI: https://doi.org/10.31289/jite.v4i2.4539.

[10] Firmansyah, F., 2021. Market Basket Analysis for Books Sales Promotion using FP Growth Algorithm, Case Study: Gramedia Matraman Jakarta. *Journal of Informatics and Telecommunication Engineering*, *4*(2), pp.383-392. DOI: https://doi.org/10.2991/assehr.k.220404.315.

[11] Aurelia, A., Dewi, N.P.L.Y., Tanupratala, W., Tsaqifah, Z. and Lina, L., 2022, April. Designing a mobile based-application using Apriori algorithm: A Business development strategy during pandemic COVID-19. In *3rd Tarumanagara International Conference on the Applications of Social Sciences and Humanities (TICASH 2021)* (pp. 1928-1935). Atlantis Press. DOI: https://doi.org/10.1109/CONIT59222.2023.10205810.

[12] Srivastava, G.R., Gera, P., Goyal, N., Sharma, A. and Rani, R., 2023, June. From Clicks to Carts: Developing an Autonomous E-Grocery Shopping System. In *2023 3rd International Conference on Intelligent Technologies (CONIT)* (pp. 1-5). IEEE. DOI: https://doi.org/10.1109/CONIT59222.2023.10205810.

[13] Koomsap, P., Tan, C.F., Lin, Y.J. and Huang, C.Y., 2021, July. Managing a Retail Store and the Associated Warehouse with a Knowledge-Driven Approach. In *International Conference on Production Research* (pp. 338-354). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-031-44373-2_20.

[14] Syahril, M., Erwansyah, K. and Yeti, M., 2020. Application of Data Mining to Determine School Equipment Sales Patterns on the Wiggle Brand Using the Apriori Algorithm. *Journal of Information Systems and Computer Systems Technology*, *3*(1), pp.118-136.

[15] Nurdin, D.A., 2015. Application of Data Mining to Analyze Sales of Goods Using the Apriori Method at the Sejahtera Lhokseumawe Supermarket. 6(1), Pp. 134–155, 2015, DOI: 10.29103/TECHSI.V7I1.184.

[16] Sintia, S., Poningsih, P., Saragih, I.S., Wanto, A. and Damanik, I.S., 2019, September. Penerapan algoritma apriori dalam memprediksi hasil penjualan sparepart pc (studi kasus: toko sentra computer). In *Prosiding Seminar Nasional Riset Information Science (SENARIS)* (Vol. 1, pp. 910-917). DOI: http://dx.doi.org/10.30645/senaris.v1i0.99.