

DETERMINATION OF SALES DATA PATTERNS USING THE ASSOCIATION RULES APRIORI METHOD

Irene Ananda¹, Umniy Salamah²

^{1,2}Departement of Computer Science, Mercu Buana
University, Jakarta, Indonesia

^{1,2}Street Meruya Selatan No. 1 Meruya Selatan,
Jakarta, Indonesia

*Corresponding author

iren.anandaw@gmail.com

umniy.salamah@mercubuana.ac.id

Article history:

Received: 28 May 2020

Revised: 29 July 2020

Accepted: 06 August 2020

Keywords:

Sales Data;

Patterns;

association rules;

Apriori

Abstract

In competition in the business world, it is necessary to find the right strategy that can be used in sales optimization. Factors that influence the needs of market analysis is the level of frequency of consumers in buying an item. Because it is needed a solution to find sales patterns with the website to be more effective and efficient. The required data is taken from sales transaction data for a certain period and processed to produce association rules for goods and transactions. Besides being able to look for patterns that often appear among many transactions, this can make it easier for companies to increase sales turnover. The making of this application uses HTML as web page development, PHP as website development, and MySQL as database management. In the testing phase, this application starts from the login to get the results of the association analysis going well. Then from the conclusion of the application made with this application the manager can add more stock of goods to the product with the highest itemset, while for the lowest itemset marketing can be done by providing a package or discount for the purchase of these items.

1.0 INTRODUCTION

Marketing strategy is a company concept that aims to achieve the sales target. In the business world, many companies try to compete to increase the company's turnover. To compete requires the right strategy in optimizing sales. This competition is caused by many companies engaged in the same field, as well as in the field of electronic distributors. Lately, Data Mining has penetrated and begun to be implemented in various fields, including in the fields of business or trade, education, and telecommunications. In business, for example, the results of applying Data Mining with the Apriori Algorithm can help business people make decisions about what is related to inventory and sales strategies. Business development is a very important factor to be taken into account. This is done to always survive in the competitive business world. Improving product quality, increasing product types, and reducing company operating costs are some things that can be done to survive in competition [1]. In making a sales strategy the factors to consider are observing the pattern of purchasing trends of an item that depends on the sale of goods or other goods. So businesses can provide the same amount of stock for each interdependent item [2].

Customer relationship management is the main goal of every business organization. In this competitive business world, every activity starts and ends with the customer. Increasing competition and a dynamic environment, each company needs to identify, anticipate, and satisfy consumers to maximize profits [3]. Consumer purchase patterns are forms of purchases made by someone or many people to get the desired item by making a purchase transaction. The master data is not only used as a company archive, but the data can also produce useful information to increase sales and product marketing strategies [4]. The number of items in a transaction that are not processed properly, because each transaction is certainly different item set owned. In the area of business transactions that occur at any time, almost all customers choose more than one type of item [5], [6]. For example in a minimarket, management needs to know the menu most sought after by consumers that can later be used by management to produce information in the form of sales predictions [7]. The management also needs to use appropriate methods to analyze consumer spending patterns in the minimarket so it can increase overall revenue or profit [8].

This research was conducted to determine the pattern of purchasing goods and combinations of items that are frequent, moderate and rarely purchased by customers based on sales transactions, which later sales transactions run effectively and the level of frequency of consumers in buying an item can be considered so that it is easy in making decisions to determine sales strategies what is suitable to apply. And find out what percentage of the results of apriori testing in improving marketing strategies.

2.0 THEORETICAL

2.1. Data Mining

Data Mining is the process of mining data or an effort made to dig up valuable and useful information on a very large database. Data mining is an analysis of a review of data sets to find unexpected relationships and summarize data in a different and understandably way that is also beneficial to the data owner. Because of that data mining has long roots from the fields of science such as artificial intelligence (artificially intelligence), machine learning, statistics, and databases [9]. Data mining is the extraction of interesting (non-trivial, implied, previously unknown, and potentially useful) patterns or knowledge from large amounts of data [10]. Data Mining has a series of processes that are divided into several stages. These steps can be illustrated in the following figure:

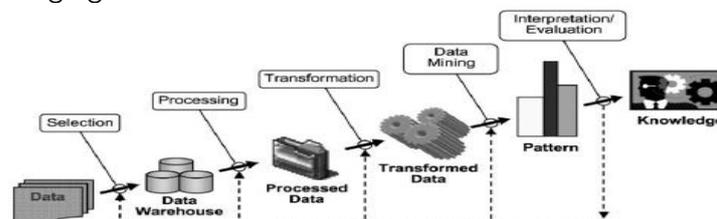


Figure 1. Data Mining Model

2.2. PHP and HTML

PHP stands for Hypertext Preprocessor which is used as server-side script language in the development of the Web inserted in the HTML document [11]. PHP is commonly used in web development that is open source. PHP and HTML are almost the same in their use to create websites, the difference is that PHP can be used to create dynamic websites, where the website can adjust the appearance of content depending on the situation. Meanwhile, HTML is usually used to create static websites, where the content or web pages are fixed.

2.3. MySQL

SQL is a special programming language that is used to access data in a relational database. Almost all database servers and software understand the SQL programming language because SQL can define tables, add data, update data, and delete data.

2.4. Website

The website is a collection of pages that display various information such as images, text, animation, and video. Websites are usually written in HTML (HyperText Markup Language) format and can be accessed via the HTTP protocol (HyperText Transfer Protocol) which will be

displayed to users through a Browser. The website is a collection of web pages that are interconnected and can be accessed through the front page (home page) using a browser [12].

2.5. Product Marketing

Product marketing is important in improving company performance. Product marketing strategy is an activity that must be carried out to introduce the product to the public. One product marketing strategy used is to increase sales optimization. One of the factors that influence product marketing is to conduct a market analysis of the level of frequency of consumers in buying an item.

3.0 METHODOLOGY

3.1. Apriori Algorithm Method

Apriori Algorithm which is one of the classic algorithms in Data Mining. Apriori Algorithm is commonly used in transaction data or can be called market basket analysis. Apriori Algorithm is used so that the computer can learn the rules of the association, looking for patterns of relationship between one or more items in the database itself. Important or not the association rules can be known by two benchmarks: support (support value) which is a percentage of a combination of items and confidence (confidence value) which is a percentage of the strength of the relationship between items in an associative rule. The process of calculating association rules consists of several steps [7]:

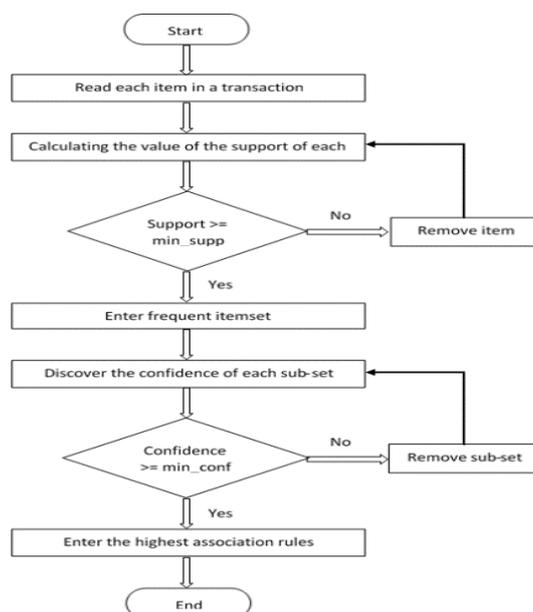


Figure 2. Flowchart of Apriori Algorithm

Analysis Calculation Algorithm Apriori:

Apriori algorithm is an algorithm that is often used to find a link between the item with other items and high frequency pattern. Apriori algorithm is divided into several stages called iterations:

1. Determining minimum support.
2. Formation candidate item set, candidate k-item set is formed from a combination of (k-1) -item set obtained from the previous iteration. One characteristic of the Apriori algorithm is the trimming candidate k-item set subset containing k-1 items is not included in the high-frequency pattern with a length of k-1.
3. The calculation of the support of each candidate k-item set. Support of each candidate k-item sets obtained by scanning a database to calculate the number of transactions that contain all the items within the candidate k-item set. It is also a characteristic of the Apriori algorithm in which the necessary calculations to scan the entire database as k-item set longest.
4. Set a high-frequency pattern. The high-frequency pattern that includes the k-item set item or set of candidate k-item set.

5. Make the process for the next iteration until no k-item set meets the minimum support [13].

3.2. Data Analysis

In this research, the input is the date, invoice number, product code, product name, and amount. While the output is in the form of the value of support to determine the association of goods sold and the value of confidence to determine the value of the association of goods that has a correlates with between items.

DATE	INVOICE	PCODE	PRODUCT NAME	QTY
03/01/2019	800001	15.4	ECF 625	1
04/01/2019	800002	15.29	EA 80 A	2
04/01/2019	800003	15.46	EWS 1118	3
04/01/2019	800003	15.32	EDB 08 B	1
07/01/2019	102/01/2019	13.18	ED 5 E	50
08/01/2019	800004	15.16	ESC 200	9
09/01/2019	800005	15.37	ECF 325	1
09/01/2019	800005	15.41	ECF 400	1
09/01/2019	103/01/2019	15.46	EWS 1118	30
10/01/2019	800006	15.45	EWS 9588	10
11/01/2019	800007	15.45	EWS 9588	10
11/01/2019	800008	15.36	EB 775 SWH	1
14/01/2019	104/01/2019	15.46	EWS 1118	5
14/01/2019	104/01/2019	15.4	ECF 625	1

Figure 3. Sales data analysis

4.0 PLANNING AND IMPLEMENTATION

4.1. System Design

The system design that will be used in this research is to use a flow map:

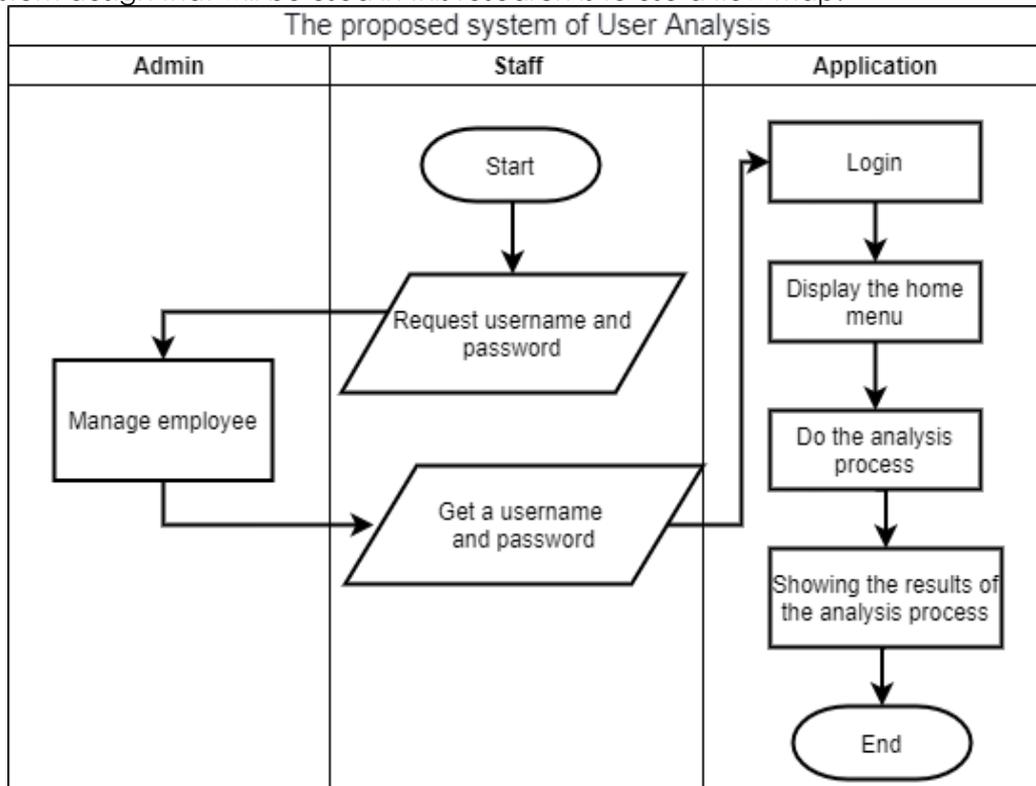


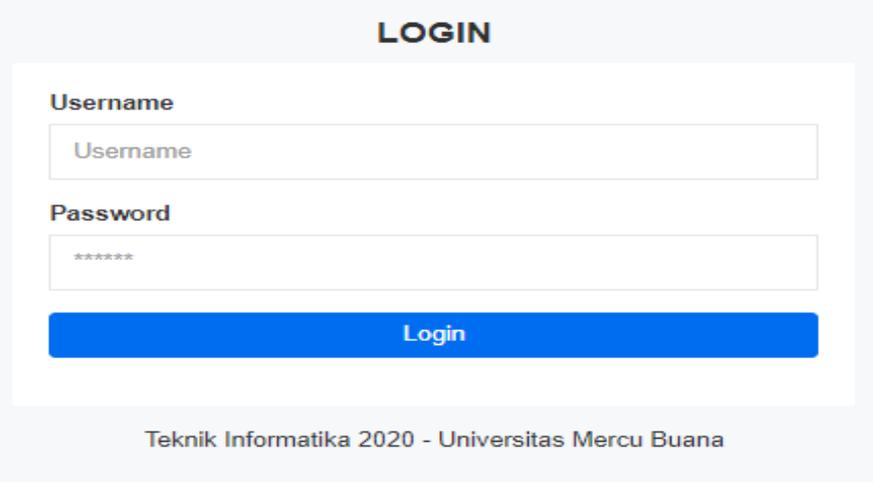
Figure 4. System design flow map

In Figure 4 above, there are the roles of admin and user. Admin functions to limit the access rights of each employee who will enter the application. Users can log in according to the username and password that has been created by the admin. If the user has logged in, the application will display the main page, then the user can carry out an analysis process to look for correlations between items.

4.2. System Implementation

Interface design is a very important part of an application. The interface is a link for users to be able to communicate with the application. The design of the interface display is needed to facilitate the process of implementing the application [14]. The use of this association application requires a username and password to maintain the security and credibility of the association application data. Display interface for logging in as shown below:

Display Login Form



The image shows a login form with the following elements:

- Title:** LOGIN
- Username Field:** A text input field with the placeholder text "Username".
- Password Field:** A text input field with masked characters "*****".
- Login Button:** A blue button with the text "Login".
- Footer:** Teknik Informatika 2020 - Universitas Mercu Buana

Figure 5. Display Login Form

The use of this association application requires a username and password to maintain the security and credibility of the association application data. Only admin with active status can log in.

Display Homepage

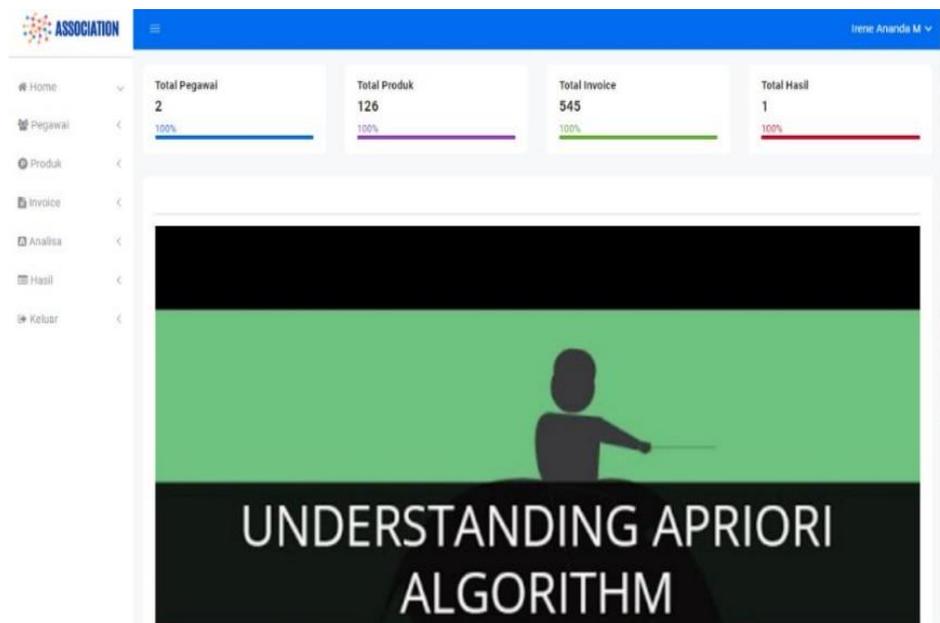


Figure 6. Display Homepage

The homepage display has 5 menus namely employee, product, invoice, analysis, and results in menus. Employee menu is used to manage user data that will use this application, manage admin data that is still active and inactive status. Product menus are used to manage product data, manage product data that is still available, and group products according to their descriptions. The invoice menu is used to manage invoice data, grouping invoices according to the date of purchase. The analysis menu is used to carry out the process of analyzing data on transaction data. And the results menu is used to see the results of the data analysis process that has been stored in the association application.

Display Analysis Menu

Figure 7. Display Analysis Menu

The analysis menu is used to carry out an analysis process in which there are two parameters of the association of percentages of support and confidence that are determined as a set value and cannot be changed, can only determine the date period to be analyzed.

Display Analysis Process

No	Itemset	Transaksi	Jumlah	Support	Confidence
6	15.36-15.51	EB 775 SWH-ED 09 TB	18	4.4226044226044	100
7	15.07-15.51	ED 03 S-ED 09 TB	18	4.4226044226044	100
8	15.42-15.51	ED 7 E-ED 09 TB	19	4.6683046683047	100
9	12.9-15.51	DFC 5015-ED 09 TB	18	4.4226044226044	100
10	15.49-15.51	EA 90 S-ED 09 TB	18	4.4226044226044	60
Total Transaksi			181		

k-itemset (k=3)
 lanjut pada tahap iterasi ketiga dengan itemset k=3 berarti kita akan membentuk kombinasi dari 3 buah itemset sebagai berikut :

No	Itemset	Transaksi	Jumlah	Support	Confidence
1	15.4-15.41-15.47	ECF 625-ECF 400-ECF 320	2	0.49140049140049	100
2	15.46-15.47-15.54	EWS 1118-ECF 320-EWS 1238	2	0.49140049140049	100
3	15.32-13.18-15.39	EDB 08 B-ED 5 E-EWS 1218	2	0.49140049140049	100
4	13.18-15.45-15.39	ED 5 E-EWS 9588-EWS 1218	2	0.49140049140049	100
5	13.18-15.39-15.42	ED 5 E-EWS 1218-ED 7 E	2	0.49140049140049	100
6	15.45-15.39-15.42	EWS 9588-EWS 1218-ED 7 E	2	0.49140049140049	100
Total Transaksi			12		

Aturan Asosiasi :
 Jika membeli ECF 625 (15.4) dan ECF 400 (15.41) punya kemungkinan akan membeli ECF 320 (15.47) = 100 %
 Jika membeli EWS 1118 (15.46) dan ECF 320 (15.47) punya kemungkinan akan membeli EWS 1238 (15.54) = 100 %
 Jika membeli EDB 08 B (15.32) dan ED 5 E (13.18) punya kemungkinan akan membeli EWS 1218 (15.39) = 100 %
 Jika membeli ED 5 E (13.18) dan EWS 9588 (15.45) punya kemungkinan akan membeli EWS 1218 (15.39) = 100 %
 Jika membeli ED 5 E (13.18) dan EWS 1218 (15.39) punya kemungkinan akan membeli ED 7 E (15.42) = 100 %
 Jika membeli EWS 9588 (15.45) dan EWS 1218 (15.39) punya kemungkinan akan membeli ED 7 E (15.42) = 100 %

Figure 8. Display Analysis Process

After determining the date period and clicking to calculate the analysis process, the application will display k-item set and association rules from transaction data.

Display Menu Results

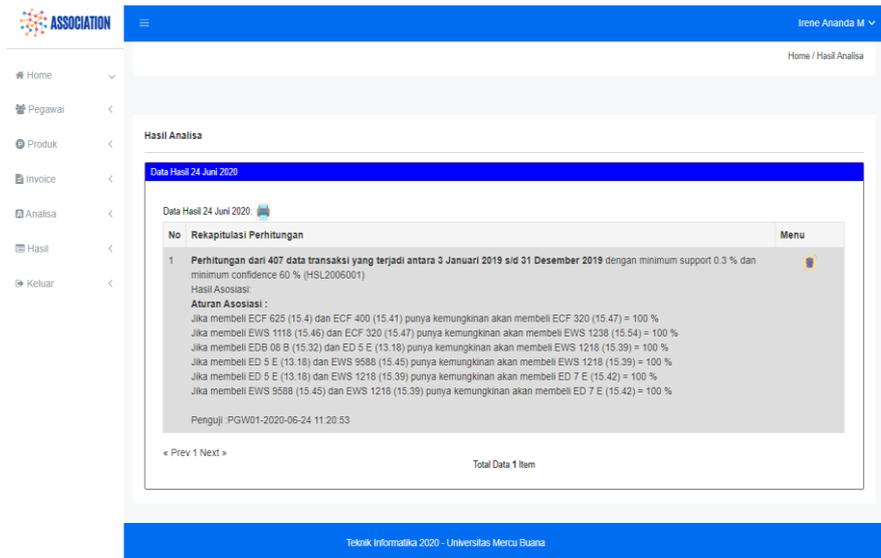


Figure 9. Display Menu Results

The results menu can display the results of the analysis that has been done by date.

4.4. Black Box Testing Scenarios

To get the appropriate results, testing the software application is required by checking directly in the application that is tailored to the application testing plan [2]. Following application testing:

1. Login Display Testing

Table 1. Login Display Testing

Input	The Results	Conclusions
Correct Username and Password	Displays notification of success	Successful
Correct Username, Incorrect Password	Failed to display a notification	Successful
Incorrect username, wrong password	Failed to display a notification	Successful
Incorrect username and password	Failed to display a notification	Successful

2. Home Display Testing

Table 2. Home Display Testing

Input	The Results	Conclusions
Notification of Success from the Login Display	Display the home view	Successful

3. Analysis Display Testing

Table 3. Analysis Display Testing

Input	The Results	Conclusions
Input date period, click calculate	Save data that has been filled	Successful
Select data and press the print button	Print the analysis results	Successful
Select data and press the save button	Save the analysis results in the results menu	Successful

4. Results Display Testing

Table 4. Results Display Testing

Input	The Results	Conclusions
Click the results data table	Showing the results of the analysis	Successful
Select data and press the delete button	Erase the analysis results	Successful
Select data and press the print button	Print the analysis results	Successful

5.0 CONCLUSION

5.1. Conclusion

With this association application, the company can overcome the problems that occur, besides that it can provide better and computerized information. The results of the web-based association application research by setting minimum support of 0.3% and a minimum trust of 60%, can produce information that the highest electronic sales in PT EPR The highest itemset that is often sought by consumers are ECF 320 Chest Freezer with product code (15.47) and items The lowest frequencies that are rarely sought by consumers, namely DFC 5015 Rice Cooker with product code (12.9), ED 03 S Water Dispenser with product code (15.07), and EA 180A Air Conditioner with product code (15.34). Then the manager can determine the marketing strategy by adding more stock of goods to the product with the highest itemset value, while for the lowest itemset value marketing can be done by providing packages or discounts for purchasing goods to attract customer interest.

5.2. Suggestion

In reference to the conclusion above, the suggestions are given to support the success of the item recommendation system is that in making this application there are still some shortcomings, especially from the interface design factor. So that the author's suggestion for further research is expected to develop the benefits of the recommendation system in various fields to produce rules that actually present more perfect item recommendations.

REFERENCES

- [1] P. B. I. S. Putra, N. P. S. M. Suryani, and S. Aryani, "Analysis of Apriori Algorithm on Sales Transactions to Arrange Placement of Goods on Minimarket," *Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 13–17, 2018.
- [2] R. Husna, R. Lestari, and Y. Hendra, "Inventory model of goods availability with apriori algorithm," *J. Phys. Conf. Ser.*, vol. 1317, no. 1, 2019, doi: 10.1088/1742-6596/1317/1/012019.
- [3] T. Kachwala and L. K. Sharma, "CSEIT1833754 | Association Rule Mining Approach for Customer Relationship Management," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3, no. 3, pp. 1991–1995, 2018.
- [4] J. Chandra and K. R. Dewi, "Implementation of Data Mining Sales of Milk Using Apriori Algorithm Method," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 662, no. 2, 2019, doi: 10.1088/1757-899X/662/2/022077.
- [5] S. Panjaitan et al., "Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012057.
- [6] Y. Sutisnawati and M. Reski, "Looking for Transaction Data Pattern Using Apriori Algorithm with Association Rule Method," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 662, no. 2, 2019, doi: 10.1088/1757-899X/662/2/022078.
- [7] Y. Nur, A. Triayudi, and I. Diana, "Implementation of Data Mining to Predict Food Sales Rate Method using Apriori," *Int. J. Comput. Appl.*, vol. 178, no. 35, pp. 22–28, 2019, doi: 10.5120/ijca2019919228.
- [8] Mustakim et al., "Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau," *J. Phys. Conf. Ser.*, vol. 1114, no. 1, 2018, doi: 10.1088/1742-6596/1114/1/012131.

- [9] Z. Wardah and D. Fitriana, "Implementasi Data Mining Pada Penjualan Tiket Pesawat Menggunakan Algoritma Apriori (Studi Kasus : PT. Pesona Ceria Travel)," *Pelita Inform. Budi Dharma*, vol. 2, no. 2, pp. 31–39, 2017.
- [10] Q. K. Kadhim, "Data Mining Concepts and Techniques مطاڪ ناعنك يصق م .م بوساحلا مولع مسق Data Mining Concepts and Techniques," no. January, 2018.
- [11] M. Muslihudin and A. Larasati, "Perancangan Sistem Aplikasi Penerimaan Mahasiswa Baru Di Stmik Pringsewu Menggunakan Php Dan Mysql," *J. TAM*, vol. 3, pp. 32–39, 2014.
- [12] A. N. Ismail, F. Sumarsono, and Nuryana, "Perancangan Website Data Karyawan Dengan menggunakan PHP dan MYSQL," *J. Sist. basis data*, no. January, pp. 1–8, 2019.
- [13] D. Sophia and L. Y. Banowosari, "Implementasi Metode Aturan Asosiasi Menggunakan Algoritma Apriori Pada Data Transaksi Penjualan Di Waroeng Spesial Sambal," *J. Inform. dan Komput.*, vol. 22, no. 1, pp. 44–56, 2017.
- [14] U. Salamah and E. Maulana, "DEVELOPMENT OF ART PERFORMANCE TICKETS INFORMATION SYSTEM AT," pp. 29–39.