Association Rules Mining for Designing Product Bundling Strategy (Study Case at a Cafe in Jakarta)

Milqi Miftahul Jannah [a,1], Yelita Anggiane Iskandar [a,2,*], Ferani Eva Zulvia [a,3]

[1,2,3] Department of Logistics Engineering, universitas pertamina ,Kebayoran Lama, Jakarta 12220, Indonesia
[1] 102417068@student.universitaspertamina.ac.id; [2] yelita.ai@universitaspertamina.ac.id *;
[3] feranieva.z@universitaspertamina.ac.id
*** corresponding author**

ARTICLE INFO

Article history
Received :24-11-2023
Revised :26-11-2023
Accepted :28-11-2023

Keywords
Data Mining;
Association Rules;
Apriori Algorithm;
Customer Relationship Management;
Product Bundling;

ABSTRACT

Cafe X is a modern and prominent coffee shop in town that sells specialty beverage products with a total of 73 variations. Sales transactions that occurred in March 2021 at that Cafe reached an average of 300 times per day. All this time, the transaction data is only used to calculate inventory replenishment and to observe sales profits whereas it can actually be utilized to reveal information related to customer behavior in buying products by finding the association or pattern between product sales. The purpose of this study is to obtain association models between items that can be used as a base to arrange recommendations for selling beverage products at Cafe X using a data mining method, namely the association rules with the Apriori algorithm. The transaction data specified for data processing are payment transactions of a total of 3,100 data records. From the analysis results, it is known that there are 8 product association rules that meet the minimum support of 0.01 and the minimum confidence of 0.5, which means that these rules can be used as an alternative basis for determining business strategy at Cafe X. Based on the association information between products, the customer relationship management (CRM) strategy that can be applied at Cafe X is the promotion strategy of product bundling. From a number of recommended bundled products, it is found that there are opportunities to increase sales for four products, they are MD001, TH002, LB001, and TU001.

1. INTRODUCTION

Coffee is one of the plantation products that have an important position in economic activity in Indonesia. Coffee also has an important role as an export product in Indonesia's foreign exchange making [1]. The Global Agricultural Information Network released the 2019 Indonesian Coffee Consumption Annual Data showing that Coffee Domestic Consumption for the 2019/2020 period of 294 thousand tons [2]. This shows an increase in coffee consumption of around 13.9% compared to the 2018/2019 period which was only 258 thousand tons. According to the International Coffee Organization (ICO), the amount of coffee consumption is more than the amount of coffee produced in the year period of 2017 to 2020 which shows that there are a lot of coffee enthusiasts in Indonesia. The increase in coffee consumption can be seen in the proliferation of cafe or coffee shop businesses with a modern concept [3]. The result of Toffin's research with Mix Magazine showed that in August 2019, the number of coffee shops in Indonesia was more than 2,950 outlets. This means that there is a threefold increase compared to 2016 when there were only 1,000 outlets [4]. The increasing number of outlets can create intense competition among similar companies [5]. Based on this, every company must find a strategy for running its business so as not to lose market share. One of the strategies that can be used to elevate the coffee business is through the implementation of Customer Relationship Management (CRM) [6]. CRM techniques can be grouped into three categories, namely strategic, operational, and analytical [7]. Analytical CRM is the process of storing, extracting, distributing, using, and reporting customer data into information that can be followed up well by the company [7].

Cafe X sales transactions in March 2021 reached an average of 300 per day. So far, the transaction data is only used to calculate inventory stock and review sales profits. In fact, the sales transaction data owned by Cafe X can also be utilized to find out a strategy to develop a business by analyzing customer satisfaction and knowing what kind of product generates the highest profit. The owner of Cafe X wants to increase its sales by utilizing the information contained in the transaction data. This study is conducted using transaction data in order to obtain information related to customer behavior in buying products by knowing the associations between products (items). Based on this, the research uses the association rule method to implement the CRM concept. The use of association rules using the Apriori algorithm in this study is aimed at gaining knowledge related to the association between items in transaction data. This knowledge is useful in generating sales strategies by adjusting the products offered to customers. Association rule is a data mining technique to find rules of relationship patterns between items in a large dataset [8]. The formed association rules can be used to determine Customer Relationship Management strategies so that Cafe X can increase product sales by providing bundled products. The parameters of the Apriopri algorithm applied in this research refer to the research developed by Syahru Romadhon and Kodar [9] which aims to form a combination of food and beverage menus using market basket analysis techniques by processing product sales transaction data. Those parameters are a minimum support of 22% and a minimum confidence of 30%. Based on the real situation at Cafe X, a problem formulation is constructed. By paying attention to customer transaction behavior and seeing the opportunity to improve business strategies, we would like to investigate the implementation of Customer Relationship Management using data mining to figure out possible recommendations to Cafe X to accelerate their business development. The aim of this research is to obtain an association model between items that are used as a basis to establish recommendations for selling beverage products at Cafe X.

2. METHOD

2.1. Data Mining

Data mining is a method of finding interesting patterns and knowledge from large amounts of data [10]. Data mining is widely used to describe and predict data [10]. An example of the application of data mining in the field of sales is the market analysis [11]. Market analysis using data mining is a model that is formed according to the concept that if a customer buys a certain item, they tend to buy another item [12]. The results of the use of data mining in market analysis can be used to find out customers' current buying patterns [12]. The use of data mining can provide information clearly and quickly by utilizing existing data to prevent errors that are detrimental to the finances of the company [13]. Data mining in CRM has 4 dimensions, namely customer identification, customer attractiveness, customer retention, and customer development, these four dimensions are useful in finding customer characteristics that are not identified in the database [14]. Data mining is used in CRM because it can save analysis costs and analyze large data quickly [7].

2.2. Association Rule

Association rule is a data mining technique for finding the rules for relationship patterns between items in a large dataset [8]. The use of the association rule using the Apriori algorithm in this study is aimed at gaining knowledge regarding the association between items in transaction data. This knowledge is expected to be used as information in making sales strategies by adjusting the products offered to customers. The results of the association rule analysis can be used to make marketing strategy policies or promotional catalog designs [15].

2.3. Customer Relationship Management

CRM is a business strategy that serves to maintain and increase the number of customers [16]. The implementation of the CRM strategy in the company is expected to create a long relationship with customers. In addition, companies can provide services that meet customer expectations that are in accordance with customer needs [6]. This study uses analytical CRM because it can store, extract, distribute, use, and report customer data into information that can be utilized positively by the company [7]. The purpose of implementing CRM is to minimize costs, namely by doing the right marketing technique to the potential customer at the right time [7].

2.4. Customer Portfolio Management

The company's customer portfolio consists of customer classification with one or more strategically crucial variables [7]. The goals given by customer portfolio management are to optimize business performance, such as increasing sales and growing profits. This can be done by providing different product offerings for different customer segments [7].

3. METHOD

3.1. Data Collection

Primary data from historical data belonging to Cafe X is sales transaction data in which there is a list of items sold, the selling price per item, and the date of the transaction, for one month in March 2021. The sales transaction considered here is explained as every sales transaction that refers to one payment transaction for Cafe X's products made by the customer, where in that one transaction, the customer buys at least 1 one beverage product.

Secondary data, namely the menu items that are sold can be seen on Cafe X's website. Table 1 contains several product names of the items sold by Cafe X which are converted into their respective item codes. In total, there are 73 product items served by the Cafe. Product codes that end in an odd number are drinks with regular sizes and those that end in an even number are drinks with large sizes except for 6 products with item codes L001 to L006 are products that have a one-liter size with 6 different flavors each.

Table 1. Product Item Code

Item Code	
SB001	SE001
SG001	VL001
CP001	HZ001
PC001	CL001
MD001	PL001
TH001	TL001
TG001	MA001
OC001	OL001
MR001	LA001
A001	CA001
KM002	FW001
HL002	SC001

From data curation, we can see that products with the item code KK001 are the most sold products. The smallest average number of sales of 73 products is 122 cups per month. Based on this, the determination of the product that sells the most is the value that has more than 122 cups. There are 9 item codes that are included in the criteria for the most sold products, namely MD001, TH002, HL001, KK001, KK002, LB001, VL001, LA001, and SC001. The criteria for the most sold products will later be used as a basic consideration for decisions to increase product sales.

Product sales data are obtained from the collection of 3,100 customer transactions within the observed period. The sales transaction taken is defined as one sales transaction referring to one payment transaction for Cafe X products made by a customer, where in one such transaction, the customer buys at least 1 drink product. The following Figure 1 is an example of a display of a sales transaction at the Cafe.

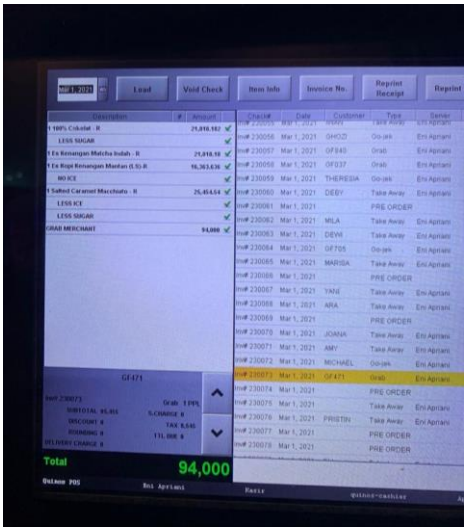

Fig. 1. Sales Transaction in the Cafe's POS

A customer portfolio is a general set of customers that includes a business's entire customer base. Based on the transaction data collected, there is some information on customers who make transactions such as the number of products purchased and how to order products. From the data collected, it is known that the number of purchases made by customers in one transaction is at least one item and at most six items. Purchase transactions with the same number of one item totaled 868 transactions and more than one item totaled 2232 transactions. This explains that 72% of 3,100 transactions made purchases of more than one item with the same product menu or different products. Product orders can be made in three ways, namely ordering using Grab, ordering using an application (pre-order), and direct ordering at the cashier (takeaway). Grab orders are orders made through online delivery services. Pre-order orders are orders made through an application from Cafe X and then use an online delivery service. Takeaway orders are orders made directly at Cafe X. The most used way for customers to order cafe products is through the Grab application and pre-orders.

Discussion

The first stage of the experiment started with working on the KDD processes, testing the minimum support and minimum confidence, determining the association rules, and analyzing the validation results including the managerial implications.

1. *KDD Stages*
   a. Data Cleaning is removing inconsistent data that occurs due to input errors or missing data. The transaction data collection results are obtained in full from Cafe X so that the data cleaning process does not need to be carried out.
   b. Data Integration is combining data sources from several data storage areas to help reduce redundancy and inconsistency in data. The data collection process at Cafe X only comes from one data storage source, namely the cash register usually known as point of sale (POS) which stores all transaction data so that the data integration process is not carried out and it is ensured that there is no repeated or redundant transaction data.
   c. Data Selection is used to reduce repetitive data and transaction data. Transactions containing purchases with the same item code only need to be written down once. For example, in one transaction where there are two purchases of products KK001 and KK001, only one item code KK001 should be written in the stored data.
   d. Data Reduction is executed to delete transactions that have only one type of item code in the stored data. This is because transactions with only one item are not affected by the purchases of other items and then do not affect the purchases of other items. This is required to narrow the solution to the association between products only.

e. Data Preprocessing is processing the data attributes needed in data mining. The attribute used in the preprocessing data in this study is the item code in one transaction. The types of data contained in sales transactions at Cafe X are at least 1 product item and at most 6 product items in every sales transaction. The transaction data used in the processing are 983 out of a total of 3,100 transactions. These 983 data are restricted to data that has the number of transactions with more than 1 item only. This study focuses on finding the relationship between items therefore data containing only 1 item is not taken into account in data processing. Examples of input data after preprocessing are shown in Table 2.

Table 2. Preprocessing Data

No.	Product					
	1	**2**	**3**	**4**	**5**	**6**
1	KK001	KK002				
2	KK001	KK002				
3	KK001	KK002				
4	MD002	MG002	KK002			
5	MD002	MG002	KK002			
6	KK001	LB001	SC002			
7	KM001	KK001	SC001			
8	MT001	KK001	KK002	LA001	SC001	
…	MD001	OC002	KK001	LA001	LA002	
983	MD001	MD002	MT001	KK001	KK002	LA001

2) Minimum Support Trial

The minimum support that will be tested is from 0.01 to 0.09 with a minimum confidence of 0.5. Table 3 displays the result of the minimum support test. We have 5 values of minimum support with 1 rule, 2 with 2 rules, and one each with 3, 4, and 8 rules.

Table 3. Recapitulation of Minimum Support Trial

Minimum Support	Minimum Confidence (0.5)
0,01	8 *rules*
0,02	4 *rules*
0,03	3 *rules*
0,04	2 *rules*
0,05	2 *rules*
0,06	1 *rule*
0,07	1 *rule*
0,08	1 *rule*
0,09	1 *rule*
0,1	1 *rule*
0,11	*no rule* (0)

3) Minimum Confidence Trial

The minimum confidence that will be tested is from 0.1 to 0.9 with a minimum support of 0.5. Table 4 shows the result of the rules from the trial of each minimum support.

Table 4. Recapitulation of Minimum Confidence Trial

Minimum Support	Minimum Confidence (0.5)
0,1	35 *rules*
0,2	26 *rules*
0,3	21 *rules*
0,4	11 *rules*
0,5	8 *rules*
0,7	2 *rules*
0,8	2 *rules*
0,9	2 *rules*

1	*no rule* (0)
0,7	2 *rules*
0,8	2 *rules*

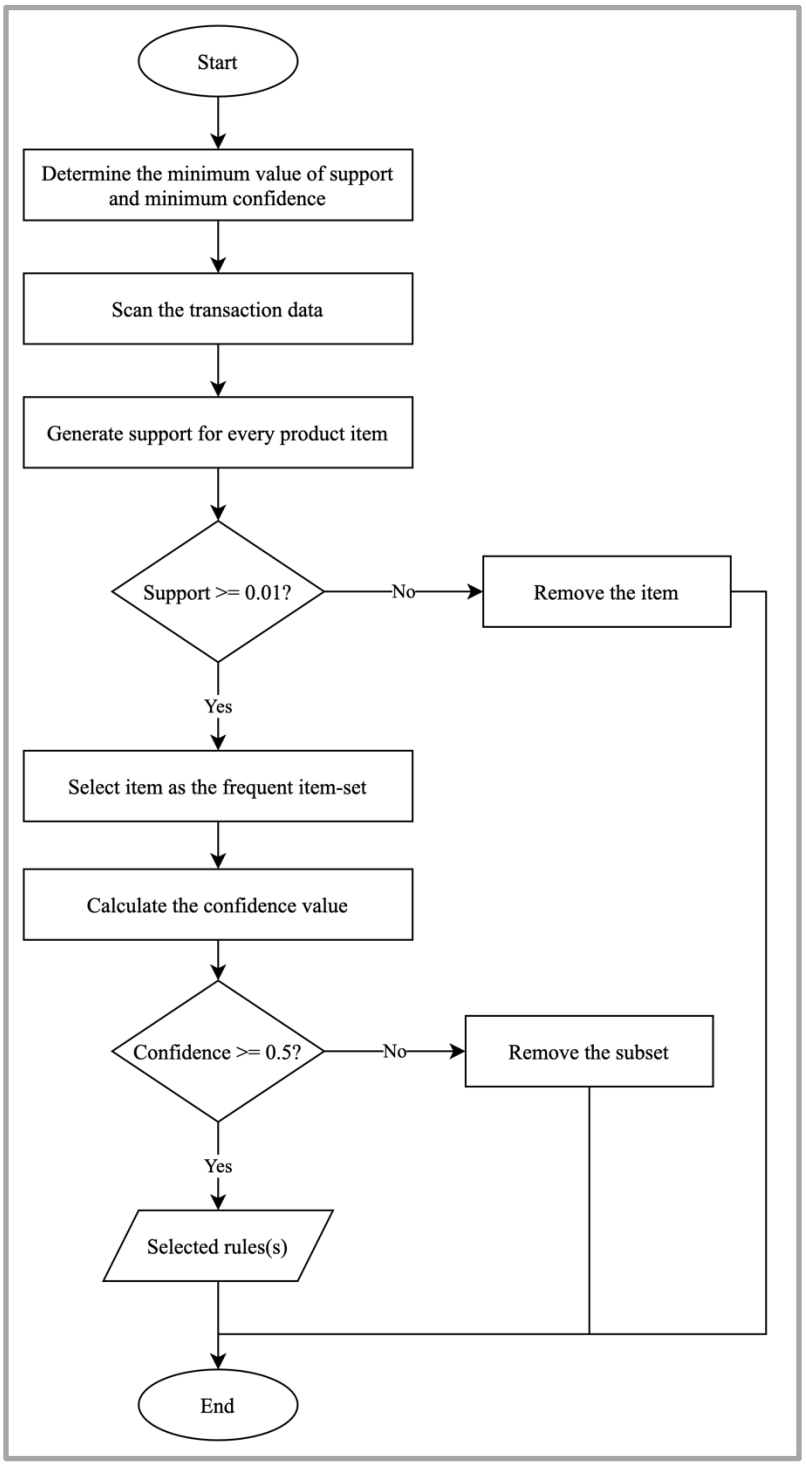

Fig. 2. Apriori Algorithm Flowchart

4) Association Rule

The algorithm used in the association rule is the Apriori algorithm built in Python. Figure 2 describes the procedure of the Apriori algorithm implementation.

Discussion

The rules that have been obtained from the processing results are in the form of code items along with their support, confidence, and lift values. These values are important information for measuring the level of association between items. Therefore the result of rules can be explained based on association rule theory in which the minimum support of 0.01 and minimum confidence of 0.5 are acknowledged as the set value here [17, 18]. The following are the final result of the rules:

1. Antecedents TU001 and consequents KK001 with support 0.01 (1%) and confidence (0.92) 90% explains that 1% of all data transactions containing items TU001 and KK001 were purchased together by customers and 90% of customers who purchased item TU001 will buy item KK001 too. The resulting lift is more than one, namely 1.88. The lift value is said to be valid based on determined support and confidence then the rules with antecedents TU001 and consequents KK001 can be used as a recommendation.
2. Antecedents L003 and consequents L001 with support 0.01 (1%) and confidence (0.91) 90% explains that 1% of all data transaction containing items L003 and L001 purchased jointly by customers and 90% of customers who purchased item L003 will also purchase item L001. The resulting lift is more than one, namely of 24.23. The lift value is said to be valid based on defined support and confidence. So the rules with antecedents L003 and consequents L001 can be used as another recommendation.
3. Antecedents LB001 and consequents KK001 with support 0.04 (4%) and confidence (0.62) 62% explains that 4% of all data transactions containing items LB001 and KK001 were purchased together by customers and 62% of customers who purchased item LB001 will also buy item KK001. The resulting lift is more than one, namely 1.28. The lift value is said to be valid based on selected support and confidence. So the rules with antecedents LB001 and consequents KK001 can be used as a recommendation.
4. Antecedents SG001 and consequents KK001 with support 0.01 (1%) and confidence (0.60) 60% explains that 1% of all data transactions containing items SG001 and KK001 were purchased together by customers and 60% of customers who purchased item SG001 will buy item KK001 too. The resulting lift is more than one, namely 1.23. The lift value is said to be valid based on determined support and confidence therefore the rules with antecedents SG001 and consequents KK001 can be used as a recommendation to be applied at the Cafe.
5. Antecedents TH002 and consequents KK001 with support 0.05 (5%) and confidence (0.58) 58% explains that 5% of all data transactions containing items TH002 and KK001 were purchased together by customers and 58% of customers who purchased item TH002 as well will buy item KK001. The resulting lift is more than one, namely 1.19. The lift value is said to be valid based on selected support and confidence. So the rules with antecedents TH002 and consequents KK001 can be recommended.
6. Antecedents TG002 and consequents KK001 with support 0.02 (2%) and confidence (0.55) 55% explains that 2% of all data transactions containing items TG002 and KK001 were purchased together by customers and 55% of customers who purchased TG002 items as well will buy item KK001. The resulting lift is bigger than one, which is 1.13. The lift value is considered to be valid based on the support and confidence picked. So the rules with antecedents TG002 and consequents KK001 are good to be applied as the Café's sales strategy.
7. Antecedents MT001 and consequents KK001 with support 0.01 (1%) and confidence (0.53) 53% explains that 1% of all data transactions containing items MT001 and KK001 were purchased together by customers and 53% of customers who purchased item MT001 will also buy item KK001. The resulting lift is greater than one which is 1.08. The lift value is said to be valid based on selected support and confidence then the rules with antecedents MT001 and consequents KK001 are great to be implemented.
8. Antecedents MD001 and consequents KK001 with support 0.10 (10%) and confidence (0.50) 50% explains that 10% of all transaction data containing items MD001 and KK001 purchased online shared by customers and 50% of customers who purchased the item MD001 will also purchase item KK001. The resulting lift is bigger than one, namely 1.02. The lift value is known to be valid based

on defined support and confidence therefore the rules with antecedents MD001 and consequents KK001 can be used as the 8th recommendation to be applied at the Cafe.

Table 5 shows the results resume of the data processing according to the flowchart in Figure 2.

Table 5. Results of Apriori Algorithm Rules

No.	If Antecedents, then Consequents	Support	Confidence	Lift
1	If buy TU001, then buy KK001	0,01	0,92	1,88
2	If you buy L003, then buy L001	0,01	0,91	24,23
3	If buy LB001, then buy KK001	0,04	0,62	1,28
4	If buy SG001, then buy KK001	0,01	0,60	1,23
5	If you buy TH002, then buy KK001	0,05	0,58	1,19
6	If buy TG002, then buy KK001	0,02	0,55	1,13
7	If buying MT001, then buy KK001	0,01	0,53	1,08
8	If you buy MD001, then buy KK001	0,10	0,50	1,02

Validation analysis is accomplished by conducting a survey to evaluate strategies arranged according to the results in Table 5. The survey was run by distributing questionnaires through online forms to loyal customers of Cafe X. The results of the survey were processed using the SPSS application, to evaluate its validity, reliability, regression, and hypotheses. Based on the survey results, it is known that respondents strongly agree with the given recommendations and choose some of the recommended bundled products. The prioritized results of the most selected products in the recommendation are shown in Figure 3.

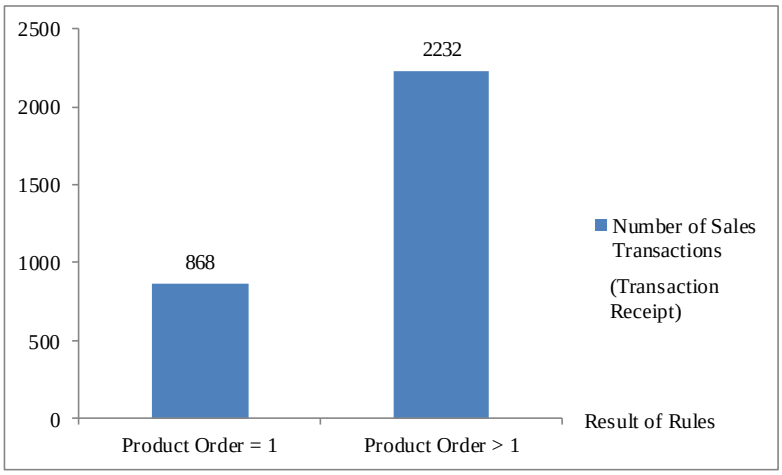

Fig. 3. Frequency of Recommended Bundled Products

The opportunity for possible sales increase can occur through the following schemes of recommended bundled products:

1. MD001 and KK001 product packages will increase MD001 sales.
2. TH002 and KK001 product packages will increase TH002 sales.
3. LB001 and KK001 product packages will increase LB001 sales.
4. TU001 and KK001 product packages will increase TU001 sales.

Based on the results to test the variables, Cafe X may implement a predetermined bundle promotion strategy. Cafe X customers are interested in bundled products because the products offered are in accordance with their preferences. Concluding from the priority of Cafe X's customers in choosing bundle product menus, Cafe X may consider the four product bundles mentioned above so that Cafe can increase its product sales.

4. CONCLUSION

Research on Cafe X has the aim of obtaining an association model between items that can be used as a basis for determining beverage product sales recommendations by taking into account the concept of customer relationship management. The association model was obtained by processing Cafe X transaction data using the Apriori data mining algorithm using the minimum support requirement of 0.01 and minimum confidence of 0.5. It is found that there are 8 association rules.

The results of the various tests of minimum support values show that the greater the value, the fewer the number of rules generated. This happened due to there are fewer items that pass the selection into the frequent itemset. Based on the results of the obtained association rules, the association between Cafe X items can be revealed so that these results can be used in making marketing strategies according to the concept of customer relationship management through offering bundled products to customers.

Recommendation validation is carried out using a customer survey, along with its validity, reliability, regression, and hypotheses test. From a number of recommended bundled products, it is known that there are opportunities to increase sales for four products, namely MD001, TH002, LB001, and TU001.

The following are some suggestions that can be used for further research development. First to use more transaction data related to data processing. This was done with the hope that the determination of the frequent itemset could be developed, thereby increasing the accuracy of the model being built. Second to conduct validation through direct implementation in the field so that the dynamics of using research results can be known in real terms.

REFERENCES

[1] Sub Direktorat Statistik Tanaman Perkebunan, *Statistik Kopi Indonesia 2019*. Badan Pusat Statistik Indonesia, 2019.

[2] United States Of Department, "Coffee: World Markets and Trade," *Foreign Agric. Serv. Off. Glob. Anal.*, 2019, [Online]. Available: http://www.fas.usda.gov/htp/coffee/2009/December_2009/2009_coffee_december.pdf.

[3] Kementrian Perindustrian Republik Indonesia, *Peluang Usaha IKM Kopi*. Jakarta: Kementerian Perindustrian Republik Indonesia, 2017.

[4] Toffin, "Riset Toffin," *tofin.id*, 2020. https://toffin.id/riset-toffin/#:~:text=Hasil riset Toffin bersama Majalah,2016 yang hanya sekitar 1.000. (accessed Apr. 22, 2021).

[5] Gobiz, "5 Tips dan Strategi Meningkatkan Penjualan Usaha Anda," 2021. https://gobiz.co.id/pusat-pengetahuan/strategi-penjualan/.

[6] A. Carissa, "Penerapan Customer Relationship Management (CRM) Sebagai Upaya untuk Meningkatkan Loyalitas pelanggan (Studi Kasus pada Bandung Sport Distro Malang)," *J. Adm. Bisnis S1 Univ. Brawijaya*, vol. 15, no. 1, p. 84471, 2014.

[7] F. Buttle and S. Maklan, *Customer Relationship Management*, 3rd ed. 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN: Routledge, 2015.

[8] M. K. Kang, "Market Basket Analysis: Identify the changing trends of market data," *Procedia Comput.*, pp. 78–85, 2016.

[9] M. SyahruRomadhon and A. Kodar, "Implementasi Metode Market Basket Analysis (Mba) Menggunakan Algoritma Apriori Dalam Transaksi Penjualan (Studi Kasus: Kafe Ruang Temu)," *J. SAINTEKOM*, vol. 10, no. 2, p. 138, 2020, doi: 10.33020/saintekom.v10i2.137.

[10] M. F. Mulya, N. Rismawati, and A. R. Rizky, "Analisis dan Implementasi Data Mining Menggunakan Algoritma Apriori untuk Meningkatkan Penjualan pada Kantin Universitas Tanri Abeng," *Fakt. Exacta*, vol. 12, no. 3, pp. 210–218, 2019, doi: 10.30998/faktorexacta.v12i3.4541.

[11] Informatika, "Data Mining Untuk Permasalahan Bisnis dan Keuangan," *FTI UNSERA*, 2017.

http://informatika.fti-unsera.id.

[12] F. Nurchalifatun, "Penerapan Metode Asosiasi Data Mining Menggunakan Algortima Apriori untuk Mengetahui Kombinasi antar Itemset pada Pondok Kopi," *Data Min.*, 2017.

[13] D. Anderson, "How Customized Data Mining Benefits Your Business," 2018. https://www.dataversity.net/ (accessed Jul. 29, 2021).

[14] N. Wardani, *Penerapan Data Mining Dalam Analytic CRM*, no. May. 2020.

[15] S. Li, "A Gentle Introduction on Market Basket Analysis — Association Rules," *towardsdatascience*, 2017. https://towardsdatascience.com/ (accessed Jul. 20, 2021).

[16] Bavarsad, "Studying the Factors Affecting the Customer Relations Management (CRM) in Maroon Petrochemical Company," *Interdiscip. J. Contemp. Res. Bus.*, 2013.

[17] C. Zhang and S. Zhang, *Association Rule Mining: Models and Algorithms*. Berlin: Springer, 2002.

[18] A. Maheshwari, *Data Analytics*. Chennai: McGraw Hill Education (India) Private Limited, 2017.