Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 3 (2023) 618 - 627 ISSN Media Electronic: 2580-0760

Analysis of Supermarket Product Purchase Transactions With the Association Data Mining Method

Norbertus Tri Suswanto Saptadi[1], Phie Chyan[2], Jeremias Mathias Leda[3]
[1,2] Department of Informatics, Faculty of Information Technology, Universitas Atma Jaya Makassar
[3] Department of Electrical Engineering, Faculty of Engineering, Universitas Atma Jaya Makassar
[1]ntsaptadi@gmail.com, [2] phiechyan@gmail.com, [3] jeremias.leda@gmail.com

Abstract

The development of business world is entering the era of big data. In meeting supermarkets' sales and purchase targets, the management needs to improve themselves in managing the goods available in the store. The research aims to determine the pattern of purchases that occur in a transaction, find out related and related products in supermarkets, and improve supermarket services to customers. The method applied uses the association rules approach to data mining. Several purchasing data from customers have been able to be analyzed by displaying a diagram as a visualization of the number of specified association rules. The processing results show a relationship above 90%: sugar and coffee with a confidence of 94.4%, shirts and trousers with a confidence of 93.4%, and sugar, milk, and coffee with a confidence of 92.0%. Decisions that can be taken by supermarket management in providing places and goods need to consider and follow product relationships and proximity based on the highest confidence value to provide services to customers effectively and efficiently.

Keywords: *transaction data' association rules; data mining; supermarket management; decisions*

1. Introduction

The corporate world is evolving towards the big data and Internet of Things era, particularly in the trade sector [1],[2]. The presence of business competition has created opportunities for competition in the business world, so every business person is expected to construct innovative, effective, and efficient ways to increase acceleration in business [3]. Supermarkets that provide goods in the world of sales are activities that aim to gain profit (profit) from consumers [4]. Business development requires governance and review of transparent and accountable values in the inventory of material stock available in warehouses sustainabl [5]. To face global business competition, every store needs to have an innovative strategy for competitive advantage [6]. Construction in material inventory management will assist in educating strategic managerial functions to predict raw materials so that they affect operations in supermarkets to be good and orderly [7].

Based on consumer needs assumptions, management must decide when to pre-order purchases from suppliers [8]. System analysis is needed to support warehouses and stores in providing measurable control in supermarket governance, such as the Economic Order Quantity method An inventory system for materials (products) in a supermarket and the types of materials that must be ready stock are needed to anticipate shortages [10]. Constraints on the stock of materials (products) that are ordered, along with sources that are difficult to access, can affect the existence of material inventory, consumer services, and supermarket revenue [11],[12]. The analysis attempted to construct a relevant system that supports decision-making so that sales optimization occurs at supermarkets and focuses on transaction data in-stock availability [13]. The analysis was carried out using a data mining approach [14],[15] on an association rule approach to obtain the best results and support governance decision-making [16].

Supermarket services have experienced an increase in the system and customer quantity. Based on data from supermarkets, the number of ingredients or producto purchases has been obtained from early 2020 to late 2021 (for two years). The sample data amounted to 1000 records from hundreds of thousands of supermarket transactions systemically. The association method in data mining is used to find the relationship between relevant products (ingredients) to help governance in decision-making [17],[18]. The discovery of patterns in purchases of various products is expected to improve services in a supermarket [19].

Accepted: 11-01-2023 | Received in revised: 02-05-2023 | Published: 02-06-2023

Based on the background stated in the introduction, it can be formulated that "how can supermarket management strategies manage and place goods to improve service to customers?". Governance determines purchasing patterns that occur in many transactions to understand the use of the association method in data mining and solve problems in supermarkets [20],[21], find out which products are interrelated and related, and improve excellent service to supermarket customers [22].

The benefits obtained are the ease of managing goods (products) in knowing purchasing patterns, layout relationships, and relationships of goods (products), providing an opportunity for management to obtain information from the results of data processing to make relevant and effective strategic decisions [23].

2. Research Methods

This research uses the associative rule method approach because it analyzes product purchases based on the relationship between one product and another so that it can obtain information regarding the layout of product positions in the store.

2.1 Transaction Data Preparation

In describing transaction patterns, 1,000 pieces of data are needed from 2020-2021. The following is a list of transactions that occur in Makassar city supermarkets.

2.2 High Frequency Form Analysis

The search for product variations on items that are the initial prerequisites for the value of support in the preparation of a database built through formula 1 [24],[25].

$$\text{Support (A)} = \frac{\text{The number of related product transactions A}}{\text{Transaction totals}} \quad (1)$$

then obtained the support value of the 2 items generated through the formula 2.

$$\text{Support } (A \cap B) = \frac{\text{The number of related product transactions A and B}}{\text{Transaction totals}} \quad (2)$$

2.3. Associative Rule Construction

All maximum frequency forms that have been obtained give an associative rule which is the minimum prerequisite for confidence. It is achieved by adding up the confidence value in the associative rules A → B. Confidence Value from rules A → B generated through the formula 3.

$$\text{Support} = P(B \mid A) = \frac{\text{The number of related product transactions A}}{\text{Transaction amount contains A}} \quad (3)$$

2.4 Association Role

I is the set which is the case study: {Coffee, Milk, …, Dolls}.

D is the set of all transactions that translate a case: {transaction 1, transaction 2, …, transaction 1000}

Proper Subset is a pure subset. Set construction A={xa,xb,xc}; A subset of A is an empty set = {}; Set in 1 Element = {xa}, {xb}, {xc}; Set in 2 elements = {xa,xb}, {xa,xc}, {xb,xc}; and the set in 3 elements = {xa,xb,xc,}; Proper subset that is formed is a set in 1 and 2 elements.

Item Set is a set of items in I, namely a set A={xa,xb,xc}; Forming item sets {xa}; {xb}, {xc}; {x,x}; {sugar}; {xb,xc}.

The K-Item Set is an Item Set that comes from K items that contain I. While K is the number of elements contained in the set: 3-item sets are 3 elements.

Item Set Frequency is the number of transactions in I which are contained in a certain number of item sets. The number of transactions that can buy an item set. Based on the supermarket transaction table, the frequency of item sets that simultaneously buy Sugar and Coffee is 4 and the frequency of item sets that simultaneously buy Coffee, Milk and Aqua is 2 (record 1-10).

Item set frequency is an item set that appears at least "a certain number of times" in D. While the word "amount" is generally symbolized by Φ, so Φ is the minimum limit in a transaction. First determine Φ = 3, because if it is not specified then the frequency of the item set cannot be calculated. If Φ=3 for {Sugar, Coffee} what is the frequency of the Item set? If it is calculated, the number of transactions that buy Sugar together with buy Coffee is 4. Because 4 >= 3 then {Sugar, Coffee} is the Frequency Item Se.

Fk is the set of all Item Set frequencies originating from K items.

2.5 Algoritma pada Association Rule

Set Φ.

Specify the entire Frequency Item set.

For each Frequency Item set has conditions: take an element with the name s; For the rest, name it ss-s; input elements that are assumed through the rule If (ss-s) then s. For the 3rd stage repeat on all elements.

3. Results and Discussions

Data mining technology uses many kinds of a priori algorithms to determine association rules. The rule defines the association between some relative attributes in affinity analysis. Another term uses market basket analysis. In analysis or association, rule mining sets combination rules on items. The a priori algorithm's goal is to identify the frequent item sets that are applied to some data to produce the form that is frequently displayed when a transaction takes place. To facilitate calculations, Rapidminer is used as an open-source software tool that is carried out in a data mining environment in the form of data mining, text mining,

DOI: https://doi.org/10.29207/resti.v7i3.4844

algorithms in machine learning, as well as in performing predictive analytics.

Prepare the data to be processed through the rapidminer studio tools before being inputted into the data processing stage. The data is processed through a priori algorithm, and the first attempted step is to determine the attributes. The attributes used or selected fulfill the relationships and relationships with one another.

RapidMiner is an open-source data science tool for performing data mining analysis, text mining and prediction analysis. Software that is used as a learning tool for data mining science, particularly when looking for links between different variables and solving problems in business operations.

The data used results from supermarket transactions totaling 1000 lines (records) with 13 attribute calculations, including sugar, coffee, milk, aqua, diapers, bed linen, washing soap, shampoo, toothpaste, toothbrush, shirts, trousers, and dolls.

3.1 Manual Calculation Process

Association rule mining with the first 10 records and 13 product items is described in transactions on items consumed and separate items purchased as shown in table 1 and 2.

Table 1. Transaction Items

Trans	Purchased Items
1	sugar, coffee, milk, aqua
2	sugar, coffee, milk, diapers, bed linen
3	diapers, bed linen, toothpaste, toothbrush, dolls
4	sugar, milk, diapers
5	sugar, coffee, aqua, shampoo, toothpaste, toothbrush
6	sugar, coffee, aqua, toothpaste, toothbrush
7	washing soap, shirts, trousers
8	aqua, shampoo, shirts
9	washing soap, shampoo, toothpaste, toothbrush, shirts, trousers
10	aqua, shampoo, toothpaste, toothbrush

Table 2. Customer Purchased Items

Trans	Item Code	Purchased Items
1	XA	Sugar
2	XB	Coffee
3	XC	Milk
4	XD	Aqua
5	XE	Diapers
6	XF	Bed linen
7	XG	Washing soap
8	XH	Shampoo
9	XI	Toothpaste
10	XJ	Toothbrush
11	XK	Shirts
12	XL	Trousers
13	XM	Dolls

Table 3, 4 and 5 are formations of transformation and purchase of each item as many as 10 and 1000 transactions.

Table 3. Transaction Transformation

Item Trans	XA	XB	XC	XD	XE	XF	XG	XH	XI	XJ	XK	XL	XM
1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	1	1	1	0	1	1	0	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	1	1	0	0	1
4	1	0	1	0	1	0	0	0	0	0	0	0	0
5	1	1	0	1	0	0	0	1	1	1	0	0	0
6	1	1	0	1	0	0	0	0	1	1	0	0	0
7	0	0	0	0	0	0	1	0	0	0	1	1	0
8	0	0	0	1	0	0	0	1	0	0	1	0	0
9	0	0	0	0	0	0	1	1	1	1	1	1	0
10	0	0	0	1	0	0	0	1	1	1	0	0	0

Table 4. Purchase Transactions 10 Records

Item Trans	XA	XB	XC	XD	XE	XF	XG	XH	XI	XJ	XK	XL	XM
1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	1	1	1	0	1	1	0	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	1	1	0	0	1
4	1	0	1	0	1	0	0	0	0	0	0	0	0
5	1	1	0	1	0	0	0	1	1	1	0	0	0
6	1	1	0	1	0	0	0	0	1	1	0	0	0
7	0	0	0	0	0	0	1	0	0	0	1	1	0
8	0	0	0	1	0	0	0	1	0	0	1	0	0
9	0	0	0	0	0	0	1	1	1	1	1	1	0
10	0	0	0	1	0	0	0	1	1	1	0	0	0
	5	4	3	5	3	2	2	4	5	5	3	2	1

Table 5. Purchase Transactions for 1000 Records

Item Trans	XA	XB	XC	XD	XE	XF	XG	XH	XI	XJ	XK	XL	XM
1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	1	1	1	0	1	1	0	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	1	1	0	0	1
4	1	0	1	0	1	0	0	0	0	0	0	0	0
5	1	1	0	1	0	0	0	1	1	1	0	0	0
6	1	1	0	1	0	0	0	0	1	1	0	0	0
7	0	0	0	0	0	0	1	0	0	0	1	1	0
8	0	0	0	1	0	0	0	1	0	0	1	0	0
9	0	0	0	0	0	0	1	1	1	1	1	1	0
10	0	0	0	1	0	0	0	1	1	1	0	0	0
	…	…	…	…	…	…	…	…	…	…	…	…	…
1000	1	1	1	1	0	0	0	0	0	0	0	0	0
Totals	478	511	475	463	388	195	392	441	369	297	335	360	291

Set Φ = 3, via an itemset frequency. From the table with 10 records, it can be obtained that the total Φ in transactions k = 1, the value is greater than Φ. So that:

F1 = {{XA}, {XB}, {XC}, {XD}, {XE}, {XF}, {XG}, {XH}, {XI}, {XJ}, {XK}, {XL}, {XM}}

k = 2 (2 elements), a table is needed for each pair of items. The set formed is:

{XA,XB}, {XA,XC}, {XA,XD}, {XA,XE}, {XA,XF}, {XA,XG}, {XA,XH}, {XA,XI}, {XA,XJ}, {XA,XK}, {XA,XL}, {XA,XM}, {XB,XC}, {XB,XD}, {XB,XE}, {XB,XF}, {XB,XG}, {XB,XH}, {XB,XI}, {XB,XJ}, {XB,XK}, {XB,XL}, {XB,XM}, {XC,XD}, {XC,XE}, {XC,XF}, {XC,XG}, {XC,XH}, {XC,XI}, {XC,XJ}, {XC,XK}, {XC,XL}, {XC,XM}, {XD,XE}, {XD,XF}, {XD,XG}, {XD,XH}, {XD,XI}, {XD,XJ}, {XD,XK}, {XD,XL}, {XD,XM}, {XE,XF}, {XE,XG}, {XE,XH}, {XE,XI}, {XE,XJ}, {XE,XK}, {XE,XL}, {XE,XM}, {XF,XG}, {XF,XH}, {XF,XI}, {XF,XJ}, {XF,XK}, {XF,XL}, {XF,XM}, {XG,XH}, {XG,XI}, {XG,XJ}, {XG,XK}, {XG,XL}, {XG,XM}, {XH,XI}, {XH,XJ}, {XH,XK}, {XH,XL}, {XH,XM}, {XI,XJ}, {XI,XK}, {XI,XL}, {XI,XM}, {XJ,XK}, {XJ,XL}, {XJ,XM}, {XK,XL}, {XK,XM}, {XL,XM}

Table 6. Candidate 2 Item Set 10 Records

T	XA	XB	f
1	1	1	P
2	1	1	P
3	0	0	S
4	1	0	S
5	1	1	P
6	1	1	P

T	XA	XC	f
1	1	1	P
2	1	1	P
3	0	0	S
4	1	1	P
5	1	0	S
6	1	0	S

T	XA	XD	f
1	1	1	P
2	1	0	S
3	0	0	S
4	1	0	S
5	1	1	P
6	1	1	P

T	XA	XE	f
1	1	0	S
2	1	1	P
3	0	1	S
4	1	1	P
5	1	0	S
6	1	0	S

T	XA	XB	f
7	0	0	S
8	0	0	S
9	0	0	S
10	0	0	S
Σ			4

T	XA	XC	f
7	0	0	S
8	0	0	S
9	0	0	S
10	0	0	P
Σ			3

T	XA	XD	f
7	0	0	S
8	0	1	S
9	0	0	S
10	0	1	S
Σ			3

T	XA	XE	f
7	0	0	S
8	0	0	S
9	0	0	S
10	0	0	S
Σ			2

T	XA	XF	f
1	1	0	S
2	1	1	P
3	0	1	S
4	1	0	S
5	1	0	S
6	1	0	S
7	0	0	S
8	0	0	S
9	0	0	S
10	0	0	S
Σ			1

T	XA	XG	f
1	1	0	S
2	1	0	S
3	0	0	S
4	1	0	S
5	1	0	S
6	1	0	S
7	0	1	S
8	0	0	S
9	0	1	S
10	0	0	S
Σ			0

T	XA	XH	f
1	1	0	S
2	1	0	S
3	0	0	S
4	1	0	S
5	1	1	P
6	1	0	S
7	0	0	S
8	0	1	S
9	0	1	S
10	0	1	S
Σ			1

T	XA	XI	f
1	1	0	S
2	1	0	S
3	0	1	S
4	1	0	S
5	1	1	P
6	1	1	P
7	0	0	S
8	0	0	S
9	0	1	S
10	0	1	S
Σ			2

T	XA	XJ	f
1	1	0	S
2	1	0	S
3	0	1	S
4	1	0	S
5	1	1	P
6	1	1	P
7	0	0	S
8	0	0	S
9	0	1	S
10	0	1	S
Σ			2

T	XA	XK	f
1	1	0	S
2	1	0	S
3	0	0	S
4	1	0	S
5	1	0	S
6	1	0	S
7	0	1	S
8	0	1	S
9	0	1	S
10	0	0	S
Σ			0

T	XA	XL	f
1	1	0	S
2	1	0	S
3	0	0	S
4	1	0	S
5	1	0	S
6	1	0	S
7	0	1	S
8	0	0	S
9	0	1	S
10	0	0	S
Σ			0

T	XA	XM	f
1	1	0	S
2	1	0	P
3	0	1	S
4	1	0	S
5	1	0	S
6	1	0	S
7	0	0	S
8	0	0	S
9	0	0	S
10	0	0	S
Σ			0

Making table 6 only on the sets {XA,XB}, {XA,XC}, {XA,XD}, {XA,XE}, {XA,XF}, {XA,XG}, {XA,XH}, {XA,XI}, {XA,XJ}, {XA,XK}, {XA,XL}, {XA,XM}, so that the entire pair table can be formed. In table 2 of the above elements, P means items that are sold together, while S means that there are no items that are sold together or there has been no transaction. Σ means the number of item set frequencies. The number of item set frequencies needs to be > or = the number of item set frequencies (Σ >= Φ). From table 6, it is obtained:

F2 = {{XA,XB}, {XA,XC}} under the condition Φ = 3.

The combination of the itemsets in F2 can be combined into a 3-itemset candidate so that the itemsets that can be combined are itemsets that have similarities in the first k-1 items. For example:

{XA,XB} and {XA,XC} have the same first k-1 itemset, namely XA, then they can be combined into a new 3-itemset, namely {XA,XB,XC}.

For k = 3 (3 elements), the possible sets are in table 7.

Table 7. Candidate Item 3 Set

T	XA	XB	XC	f
1	1	1	1	P
2	1	1	1	P
3	0	0	0	S
4	1	0	1	S
5	1	1	0	S
6	1	1	0	S
7	0	0	0	S
8	0	0	0	S
9	0	0	0	S
10	0	0	0	S
Σ				2

From table 7, we get F3 = {XA,XB,XC}, because there is no Σ = Φ.

The Rules used are if x then y, where x is the antecedent and y is the consequent. Based on this rule, 2 items are needed, one of which is the antecedent and the rest is the consequent. From step 5, we get 1 Fk, namely F2. F1 is not included because it only consists of 1 item. The antecedent may have more than 1 element, while the consequent consists of 1 element.

Determine (ss-s) as the antecedent and s as the consequent of the Fk that has been obtained based on the rule in step 6.

In F2 we get the set F2 = {{XA,XB}, {XA,XC}, then it can be arranged:

To {XA,XB}:
If (ss-s) = XA, If s = XB, Then If buy XA then buy XB
If (ss-s) = XB, If s = XA, Then If buy XB then buy XA
To {A,C}:
If (ss-s) = XA, If s = XC, Then If buy XA then buy XC
If (ss-s) = XC, If s = XA, Then If buy XC then buy XA

From step 7, you will get 4 rules that can be used.

If buy XA then buy XB
If buy XB then buy XA
If buy XA then buy XC
If buy XC then buy XA

From step 8, table 8, a candidate association rules for 1 antecedent is made.

Table 8. Candidate Item 3 Set

If antecedent then consequent	Support	Confidence
If buy XA then buy XB	………………..	……………….
If buy XB then buy XA	………………..	……………….
If buy XA then buy XC	………………..	……………….
If buy XC then buy XA	………………..	……………….

Formula 4 and 5 are used to calculate support and confidence.

$$\text{Support (A)} = \frac{\Sigma \text{ Items purchased together}}{\Sigma \text{ The totals number of transaction}} \quad (4)$$

$$\text{Support (B)} = \frac{\Sigma \text{ Items purchased together}}{\Sigma \text{ Number of antecedent transactions}} \quad (5)$$

Σ items that are simultaneously purchased If buy XA then buy XB there are 4 transactions out of 10 transactions so that the support (assuming 10 transactions/record of 1000 transactions/record) is shown in formula 6.

$$\text{Support} = \frac{4}{10} x\ 100\% = 40\% \quad (6)$$

DOI: https://doi.org/10.29207/resti.v7i3.4844

For Σ items purchased at once on If buy XA then buy XB, there are 4 transactions, while the number of transactions that buy XA is 5 transactions, so the confidence is shown in formula 7.

$$\text{Support} = \frac{4}{5} x\ 100\% = 80\% \tag{7}$$

So that table 9 can be obtained.

ExampleSet (1001 examples, 0 special attributes, 15 regular attributes) Filter (1,001

Row No.	Customer	Tanggal	Sugar	Coffee	Milk	Aqua	Diaspers	Bed Linen	Washing Soap	Shampoo	Toothpaste	Toothbrush	Shirts	Thousers	Dolls
1	1	Feb 3, 2020	1	1	1	1	0	0	0	0	0	0	0	0	0
2	2	Feb 3, 2020	1	1	1	0	1	1	0	0	0	0	0	0	0
3	3	Feb 3, 2020	0	0	0	0	1	1	0	0	1	1	0	0	1
4	4	Feb 3, 2020	1	0	1	0	1	0	0	0	0	0	0	0	0
5	5	Feb 3, 2020	1	1	0	1	0	0	0	1	1	1	0	0	0
6	6	Feb 3, 2020	1	1	0	1	0	0	0	0	1	1	0	0	0
7	7	Feb 3, 2020	0	0	0	0	0	0	1	0	0	0	1	1	0
8	8	Feb 3, 2020	0	0	0	1	0	0	0	1	0	0	1	0	0
9	9	Feb 5, 2020	0	0	0	0	0	0	1	1	1	1	1	1	0
10	10	Feb 5, 2020	0	0	0	1	0	0	0	1	1	1	0	0	0
11	11	Feb 5, 2020	0	0	1	0	1	0	0	0	0	0	0	0	0
12	12	Feb 5, 2020	0	0	0	0	0	1	1	1	0	1	1	1	0
13	13	Feb 5, 2020	1	1	1	1	1	0	0	0	0	0	0	0	0
14	14	Feb 5, 2020	0	0	0	0	0	0	1	1	0	0	1	1	1
15	15	Feb 5, 2020	0	0	0	0	1	1	1	0	0	0	0	0	0
16	16	Feb 5, 2020	0	0	1	0	1	1	0	1	0	0	0	0	1
17	17	Feb 5, 2020	1	1	1	1	0	0	0	0	0	0	0	0	0
18	18	Feb 5, 2020	0	1	1	1	0	0	0	0	0	0	0	0	0
19	19	Feb 5, 2020	0	0	0	0	0	0	1	0	0	0	1	1	0
20	20	Feb 5, 2020	0	0	0	0	0	0	1	0	0	0	1	1	0

Figure 1. Transaction Value Conversion Data

Table 9. Support and Confidence

If antecedent then consequent	Support	Confidence
If buy XA then buy XB	(4/10) x 100% = 40%	(5/5) x 100% = 100%
If buy XB then buy XA	(4/10) x 100% = 40%	(4/5) x 100% = 80%
If buy XA then buy XC	(3/10) x 100% = 33,3%	(5/5) x 100% = 100%
If buy XC then buy XA	(3/10) x 100% = 33,3%	(3/5) x 100% = 60%

After the support and confidence values are obtained for each candidate, the multiplication between support and confidence (taken > 70%) is measured, so that table 10 is obtained.

Table 10. Support and Confidence

If antecedent then consequent	Support	Confi-dence	Support x Confidence
If buy XA then buy XB	(4/10) x 100% = 40%	100%	0,4000
If buy XB then buy XA	(4/10) x 100% = 40%	80%	0,3200
If buy XA then buy XC	(3/10) x 100% = 33,3%	100%	0,3330

After the multiplication of support and confidence results are obtained, determine the biggest value, which is the rule used when selling. The multiplication results of the 3 sales have different values, so it can be used as a rule to make a relationship among sugar, coffee, and milk. The following is the formulation of the multiplication results that have been produced.

If you buy XA, you will buy XB with 40% support and 100% confidence.

If you buy XB, you will buy XA with 40% support and 80% confidence.

If you buy XA, you will buy XC with 33.33% support and 100% confidence.

3.2 Implementation

To be able to find out in detail, a digital application can help analyze and find out exactly, precisely, and according to the purchase transaction of a supermarket product in the city of Makassar, South Sulawesi, Indonesia as shown in figure 1.

Utilization of Rapid Miner, the company's supported platform provides integrated environments for data preparation, machine/deep learning, text mining, and predictive analytics. The process and adding operators in data mining using RapidMiner Studio. Determination of attributes through the data preprocessing stage. Perform data cleaning and data transformation to clean sales transaction data in supermarkets.

To expedite the outcomes of a combined pattern for data transformation, the cleaning stage is completed by deleting incomplete or empty data, followed by unneeded fields. It aims to make it easier when entering data into the rapidminer tool by making initials or symbols for data.

In the created rapid miner process, the data is filtered and transformed into tabular data so that it becomes data prepared to be processed in the data mining system process using the a priori algorithm. From the 10,000 data records, after filtering the data that was ready to use, there were 1,000 data records with attributes in the excel format used: Customer ID (customer id) and Description (item name).

The initial dataset has several attributes and is modified through the process. The coding transformation stage with data processing totaling 1000 records that enters data into the Rapid Miner application, then the system processes the a priori algorithm using the association rules method to get the final results as expected. Import data from excel (xlsx) to Rapid Miner Studio in binary form to true and false (numbers 1 and 0) with 1000 records and 13 regular attributes.

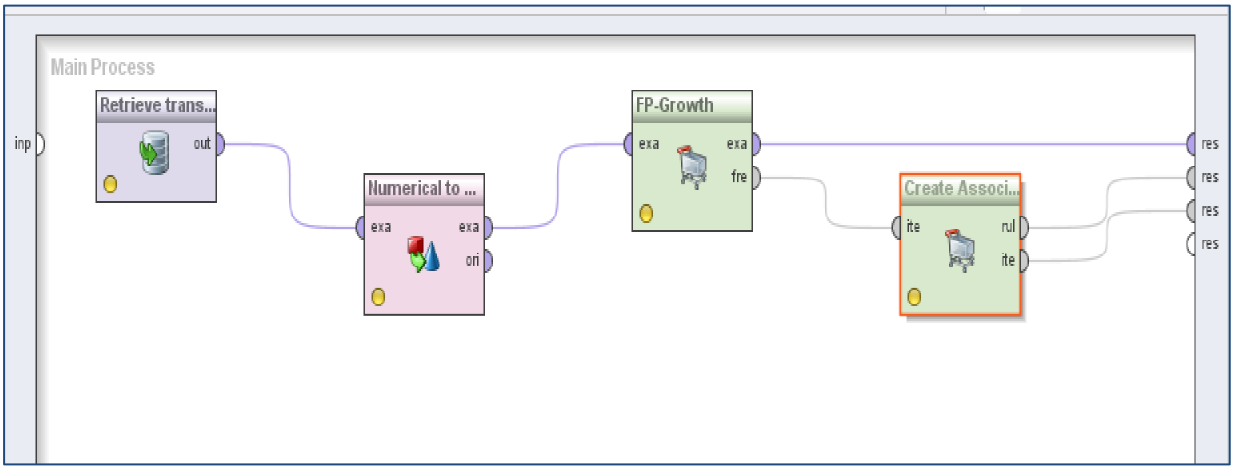


Figure 2. Process Design

Then make a process design as seen in figure 2 by retrieving 1000 transactions, numerical to binomial, fp-growth, and create association rules. The following steps are taken, namely:

Retrieve Transaction 1000, means to import data from local Repository.

The operator "Numerical to Binomial" changes the type of a numeric attribute to a binomial type. This attribute value corresponds to a certain binomial value. A numeric attribute's type is changed to a binomial (binary) type with the Numerical to Binominal operator. The operator overrides the specified attribute type but maps various attribute values to the relevant binomial values. The binomial attribute can have two values, namely 'true' or 'false'. If the attribute value is between the specified minimum and maximum values it will be 'false', otherwise 'true'. The min and max parameters can ascertain each minimum and maximum value. The new value will likewise be lost if the original value is. Both have a default limit of 0.0. Just 0.0 is default set to "false," whereas all other values are mapped to "true."

FP-Growth, means that the Operator makes efficient efforts to enumerate all frequent item sets from the ExampleSet by using the FP-tree data structure. All ExampleSet input attributes must be binomial. Frequent itemsets are a set of items that frequently appear together in the data. It is useful to know the basics of market basket analysis, which is a collection of items that occur often. The data market basket model describes the general form of a many-to-many relationship between two particular types of objects. On the other hand, the other side with baskets becomes a 'transaction' status by owning goods. The set of items is represented as a set of attributes. Most of these attributes are binomial.

Create Association Rules means Operator will generate a set of association rules from the specified set of frequent itemsets. Association rules are if/then statements that help generate relationships between mutually unrelated data. An example of an attribution rule is "If a customer buys a shirt, he is 80% likely also to buy trousers." Association rules have a status of antecedents (if) and consequences (then). Antecedents are items (or itemsets) obtained in the data. Consequences are items (or itemsets) obtained in combination with antecedents. Association rules are created by analyzing data for "if/then" patterns and using support and confidence criteria to identify the most relevant relationships. Support indicates how likely an item is to appear in the database. Confidence sets the number of times an if/then statement is proven true. Forms are often if/then mined using operators such as the FP-Growth operator. The things that are most likely to result in an association rule are obtained by the Create Association Rules operator. Making decisions on marketing efforts, such as setting promotional prices or choosing products, can be done using this information. Numerous industries still employ the market basket analysis association rule, including bioinformatics, web mining, intrusion detection, and intrusion prevention.

Application of ExampleSet (Numerical to Biominal), displays data from the results of importing data from the local Repository can be seen in figure 5.

Displays statistics for each of the 13 attributes, namely: sugar, coffee, milk, aqua, diapers, bed linen, washing

DOI: https://doi.org/10.29207/resti.v7i3.4844

soap, shampoo, toothpaste, toothbrush, shirts, trousers, and dolls. Statistics display data type (type), empty data (miss), minimum value condition (min), maximum value condition (max), average value result (average), and standard deviation. For data types using integers, miss is 0, min is 0, max is 1. The highest average of 0.478 is in the sugar attribute, and the lowest average of 0.195 is found in the bed sheet attribute. The highest standard deviation is 0.5 on the attributes of sugar, coffee, milk and the lowest is 0.396 on bed linen.

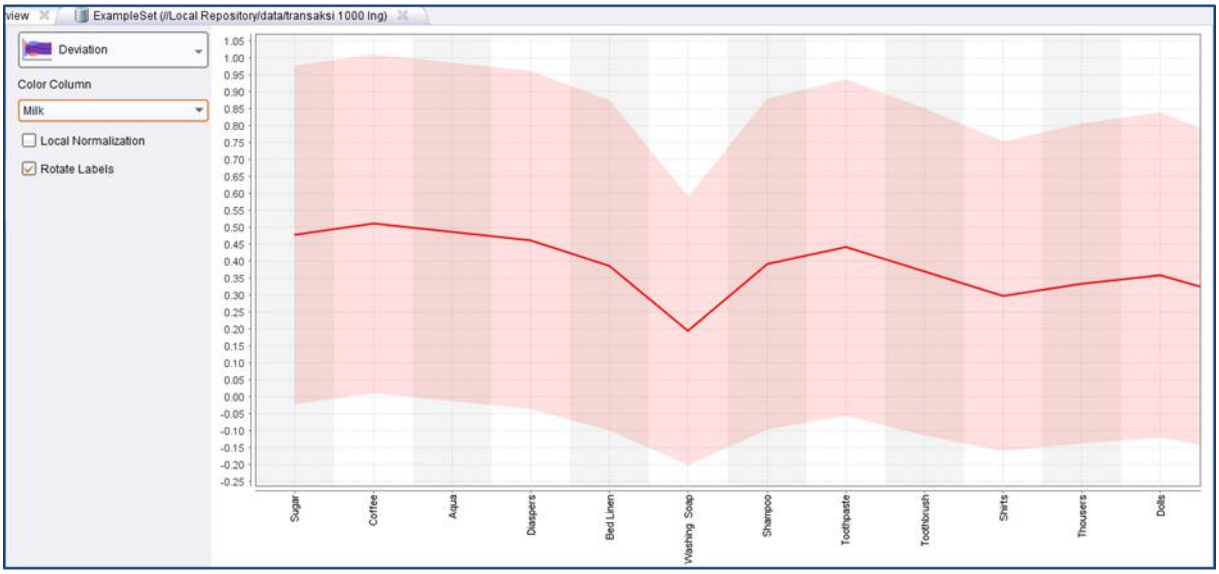


Figure 3. Chart Destination

Scatter Multiple, the chart in figure 4 display shows that there is a multiple Scatter options. The x-Axis selected is sugar and the y-Axis includes sugar, coffee, milk, aqua, diapers, bed linen, washing soap, shampoo, toothpaste, toothbrush, shirts, trousers, and dolls (13 attributes) with the jitter setting with a range of 1.6 to -0.6.

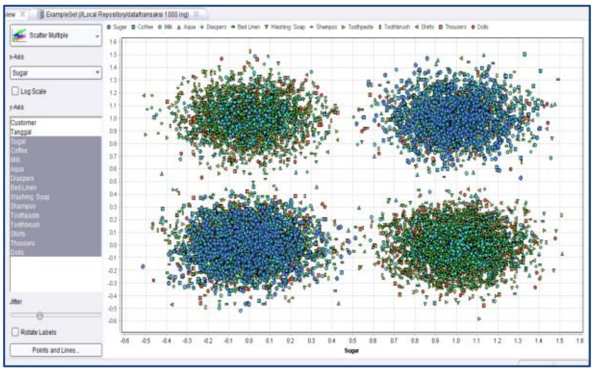


Figure 4. Scatter Multiple 1

In addition, there is a different view by selecting a jitter range of 2.75 to -2.00 with rotate labels. The number of attributes selected y-Axis is 13 attributes.

Scatter Multiple, the chart in figure 6 on scatter multiple choosing color column in aqua attribute with rotate labels. The highest value fluctuation is 0.5 on the milk attribute and the lowest is 0.2 on the washing soap attribute.

Figure 5. Scatter Multiple 2

Frequent Item Sets (FP-Growth) – Data, for FP-Growth's display of data, there are groupings (figure 6), namely:

Size 1 has the highest value of 0.511 on the coffee attribute and the lowest is 0.291 on the doll attribute.

Size 2 has the highest score of 0.451 on the attribute coffee - sugar and the lowest is 0.218 on the attribute washing soap-trousers.

Size 3 value is 0.309 on the attribute coffee-sugar-milk.

Association Rules (Create Association Rules), for the data results of the association rules, there are 7 (seven) rules formed. The rule includes the highest premises-conclusion value between the coffee-milk attributes with a support of 0.451, a confidence of 0.944. After that premises-conclusion on the shirt-trousers attribute

DOI: https://doi.org/10.29207/resti.v7i3.4844

with support 0.313 confidence 0.934. On the other hand, there is a premises-conclusion value between sugar-milk-coffee attributes with support 0.309, confidence 0.920.

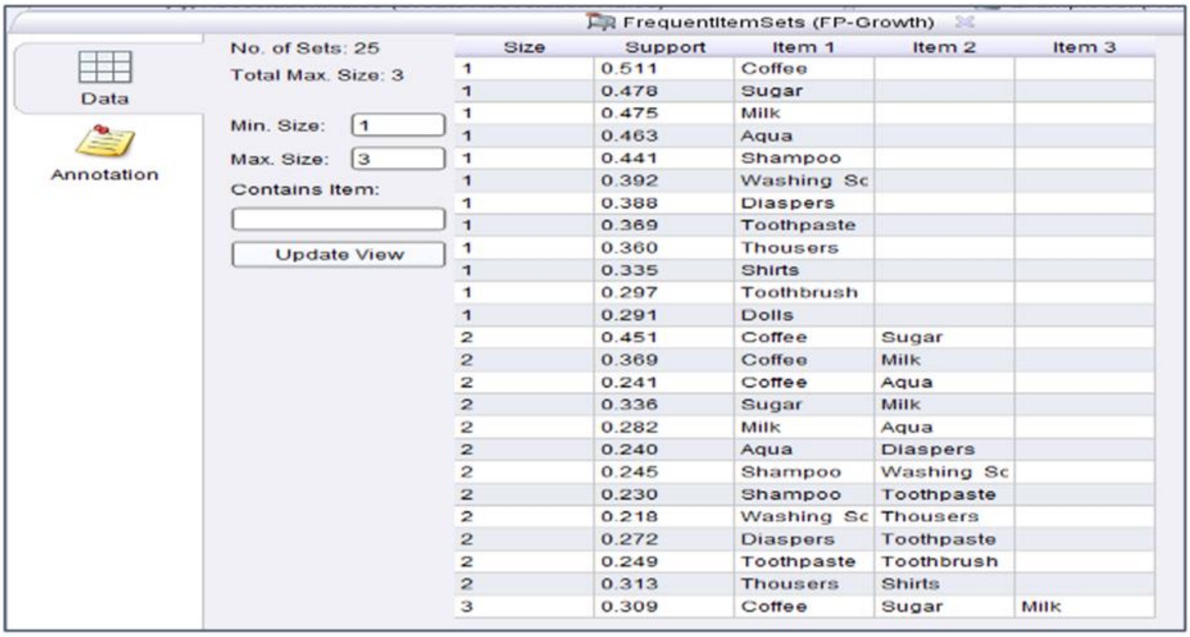
Graph, to find out the association between attributes, the following graph plot is shown:

Association of 1 (first) attribute of milk, coffee, and sugar with rules 1, 4, 5, and 7.
Association of 2 (second) attributes of toothpaste and toothbrush with rule 2.
Association of 3 (third) attributes of trousers and shirts with rules 3 and 6.

Association Rules, are :

[coffee, sugar] → [sugar] (confidence: 0,837).
[toothbrush] → [toothpaste] (confidence: 0,838).
[trousers] → [shirt] (confidence: 0,869).
[coffee] → [sugar] (confidence: 0,883).
[sugar, milk] → [coffee] (confidence: 0,920).
[shirt] → [trousers] (confidence: 0,934).
[sugar] → [coffee] (confidence: 0,944).

Description for the final result of supermarket data processing, which amounts to 10,000 records, 7 (seven) rules have been formed. The first highest value is in the attributes of sugar and coffee, with a confidence value of 0.944. The second is in the attributes of shirts and trousers, with a confidence value of 0.934. Moreover, the third is in the attributes of sugar, milk, and coffee, with a confidence value of 0.920.



Size	Support	Item 1	Item 2	Item 3
1	0.511	Coffee		
1	0.478	Sugar		
1	0.475	Milk		
1	0.463	Aqua		
1	0.441	Shampoo		
1	0.392	Washing Sc		
1	0.388	Diaspers		
1	0.369	Toothpaste		
1	0.360	Thousers		
1	0.335	Shirts		
1	0.297	Toothbrush		
1	0.291	Dolls		
2	0.451	Coffee	Sugar	
2	0.369	Coffee	Milk	
2	0.241	Coffee	Aqua	
2	0.336	Sugar	Milk	
2	0.282	Milk	Aqua	
2	0.240	Aqua	Diaspers	
2	0.245	Shampoo	Washing Sc	
2	0.230	Shampoo	Toothpaste	
2	0.218	Washing Sc	Thousers	
2	0.272	Diaspers	Toothpaste	
2	0.249	Toothpaste	Toothbrush	
2	0.313	Thousers	Shirts	
3	0.309	Coffee	Sugar	Milk

Figure 6. Frequent Item Sets

The results of the Association Rules show that there is a relatively strong relationship between one product and another as shown in the results of data processing. Rapid Miner is used to find hidden patterns or previously unknown association rules from a set of supermarket data. A number of purchase data from customers have been able to be analyzed by displaying a diagram as a visualization of the number of specified association rules.

Through various approaches in data mining that are grouped or collected such as association analysis, data mining is very useful in helping human life and business. The use of data mining is able to solve business problems and strengthen management. Managers can be assisted in making appropriate, fast and strategic decisions. Through good goods management, it will assist management in developing the business and improving the quality of service to consumers so that they can achieve satisfaction in shopping and meet mutual expectations. Especially to improve service to consumers.

The association method is very appropriate for finding what products are usually sold together and finding out what rules cause these similarities to occur. In order to improve store service performance, the management in particular is expected to continue to pay attention to, take into account and define each customer's needs through an associative analysis approach.

Association Rule is applied to read consumer purchasing patterns or tendencies. This study took a sample of 1000 transactions where there were several large transactions made by consumers with relatively the same pattern. For example, consumers will buy sugar if they also buy coffee, buy toothbrushes if they buy toothpaste, buy trousers if they buy shirts and so on. This is of course in line with the purpose of association rules which make it possible to obtain information about what items/goods consumers often buy together at the same time.

Information regarding sales results will be used by store management to make a policy such as management being able to replace goods/items that are often purchased together so that consumers can easily see and access them. Consumer needs for a product will be fulfilled by setting the layout of goods that are good and orderly.

A case study of a store in Makassar regarding a product of coffee and sugar, toothbrushes and toothpaste, trousers and shirts which are often purchased together will enable the management to arrange coffee and sugar placements close together to make it easier for consumers to buy both. The same goes for certain other products. Management can carry out promotions for products that are rarely purchased to reduce stock that has accumulated in the warehouse. Another thing is related to stock management, where management can make decisions to increase the stock of goods/items that are often purchased so that there is no buildup in stores or warehouses.

Association rule production is inextricably linked to association analysis, which employs the causal concept of antecedent-consequent as a key component. An technique to data exploration known as the Association Rule is frequently used or applied to very large data sets. The association rules at least offer useful data for determining if two sets of data are significantly correlated.

The Apriori algorithm, a type of association rule in data mining, was the algorithm employed in this investigation. The term "affinity analysis" or "market basket analysis" refers to rules that specify the relationship between several attributes. A data mining technique for determining the rules for a combination of elements is association analysis or association rule mining. Research will be done to create an effective algorithm with high-frequency pattern analysis (frequent pattern mining), starting with the attention-grabbing association analysis stage. Two metrics, called support and confidence, can be used to determine an association's significance. The outcomes of data processing have received confidence (certainty value) in the form of strong correlations between the findings and support (support value) in the form of a percentage combination of items in the database.

4. Conclusion

Based on manual calculations and using Rapid Miner Studio, a description of the relationship (association) between one product and another can be obtained. Of the 13 products in 1000 transactions, it is known that there is a relationship between one variable or several variables. From here it can be seen various relationships between these variables. The associative rule of purchasing analysis in a supermarket is to find out the management in determining how likely it is that a customer (customer) buys sugar together with coffee and so on so that management can arrange products or goods to be placed in the supermarket. Several purchasing data from customers have been able to be analyzed by displaying a diagram as a visualization of the number of specified association rules. The processing results show a relationship above 90%: Sugar and Coffee with a confidence of 94.4%, T-shirts and trousers with a confidence of 93.4%, and Sugar, Milk and Coffee with a confidence of 92.0%. The supermarket governance makes decisions on where to put things and how close things are to each other based on the highest confidence value in order to serve customers effectively, efficiently, and appropriately.

Acknowledgment

Thank you to the Institute for Research and Community Service, University of Atma Jaya Makassar, and Supermarkets in Makassar City who provided support in the form of licensing and in financial terms.

Reference

[1] N. Bessis and C. Dobre, *Big Data and Internet of Things : A Roadmap for Smart Environments*.

[2] K. Nagorny, P. Lima-monteiro, J. Barata, and A. W. Colombo, "Big Data Analysis in Smart Manufacturing : A Review," pp. 31–58, 2017, doi: 10.4236/ijcns.2017.103003.

[3] V. Pitoy, A. Tumbel, and M. Tielung, "Competitive Strategy Analysis in Document Solution Business Competition (Case Study at Pt. Astragraphia, Tbk Manado)," *Effic. Sci. Period. J.*, vol. 16, no. 3, pp. 302–312, 2016.

[4] N. F. Naini, Sugeng Santoso, T. S. Andriani, U. G. Claudia, and Nurfadillah, "The Effect of Product Quality, Service Quality, Customer Satisfaction on Customer Loyalty," *J. Consum. Sci.*, vol. 7, no. 1, pp. 34–50, 2022, doi: 10.29244/jcs.7.1.34-50.

[5] C. Ortega-Rodríguez, A. Licerán-Gutiérrez, and A. L. Moreno-Albarracín, "Transparency as a key element in accountability in non-profit organizations: A systematic literature review," *Sustain.*, vol. 12, no. 14, 2020, doi: 10.3390/su12145834.

[6] T. Suhaeni, "The Effect of Innovation Strategy on Competitive Advantage in Creative Industries (Case Study of MSMEs in Handicrafts in Bandung City)," *J. Bus. Invest. Res.*, vol. 4, no. 1, p. 57, 2018, doi: 10.35697/jrbi.v4i1.992.

[7] Y. Hernawati, N. P. Mulyadi, T. R. Lestari, and D. Faidz, "Evaluation of Finished Goods Stock Control System in Pt.Indocare Citra Pasifik Group Warehouse," vol. 6, pp. 20–27, 2020.

[8] Tiris Sudrartono, "The Effect of Pre Order Online Sales to Company Profits on PT. Aventama Hervent Solusindo Bandung," *J. Sos. Sci.*, vol. 1, no. 3, pp. 88–92, 2020, doi: 10.46799/jsss.v1i3.30.

[9] C. A. Puspadev, "Raw Material Inventory Control System Using the Economic Order Quantity Method at Pizzahut Setiabudi," *Informatics Sci. J.*, vol. 9, no. 01, pp. 43–47, 2021, doi: 10.33884/jif.v9i01.3725.

[10] D. N. Rorim Panday, "Inventory Management Evaluation And Inventory Forecast Using EOQ," *Syntax Lit.*, vol. 6, no. 1, p. 6, 2021.

[11] L. Sholeha, S. Djaja, and J. Widodo, "The Influence of Service Quality on Customer Satisfaction at Ahass Sumber Jaya Maha Sakti, Rogojampi District, Banyuwangi Regency," *Econ.*

DOI: https://doi.org/10.29207/resti.v7i3.4844

Educ. J. Sci. J. Educ. Econ. Soc. Sci., vol. 12, no. 1, p. 15, 2018, doi: 10.19184/jpe.v12i1.6465.

[12] Taupik Ismail, “The Effect of Service Quality on Customer Satisfaction at the Indihome Gegerkalong Office in Bandung City,” *MEA Sci. J. (Management, Econ. Accounting)*, vol. 5, no. 1, pp. 1124–1135, 2021.

[13] W. Reinartz, N. Wiegand, and M. Imschloss, “The impact of digital transformation on the retailing value chain,” *Int. J. Res. Mark.*, vol. 36, no. 3, pp. 350–366, 2019, doi: 10.1016/j.ijresmar.2018.12.002.

[14] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.

[15] Parteek Bhatia, *Data mining and data warehousing*, vol. 47. 2019.

[16] R. Turner, “How Does Governance Influence Decision Making on Projects and in Project-Based Organizations?,” *Proj. Manag. J.*, vol. 51, no. 6, pp. 670–684, 2020, doi: 10.1177/8756972820939769.

[17] M. Supriyamenon and P. Rajarajeswari, “A review on association rule mining techniques with respect to their privacy preserving capabilities,” *Int. J. Appl. Eng. Res.*, vol. 12, no. 24, pp. 15484–15488, 2017.

[18] T. Husain and N. Hidayati, “The Optimize of Association Rule Method for the Best Book Placement Patterns in Library: a Monthly Trial,” *Teknokom*, vol. 4, no. 2, pp. 53–59, 2021, doi: 10.31943/teknokom.v4i2.63.

[19] X. Chen, B. Kassas, and Z. Gao, “Impulsive purchasing in grocery shopping: Do the shopping companions matter?,” *J. Retail. Consum. Serv.*, vol. 60, no. February, p. 102495, 2021, doi: 10.1016/j.jretconser.2021.102495.

[20] S. C. Hui and G. Jha, “Data mining for customer service support,” *Inf. Manag.*, vol. 38, no. 1, pp. 1–13, 2000, doi: 10.1016/S0378-7206(00)00051-3.

[21] A. Chailes, A. Hermawan, and D. Kurnaedi, “Application of the Data Mining Method to Determine Purchasing Patterns Using the Apriori and Fp-Growth Algorithms at Mukara Stores,” *J. Algor*, vol. 1, no. 2, pp. 1–8, 2020.

[22] T. Desara and T. Hidayat, “Enrichment: Journal of Management is Licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0) Enrichment: Journal of Management Enhancing Repurchase Intention in Retail: the Role of Customer Satisfaction, Servic,” *Enrich. J. Manag.*, vol. 12, no. 1, pp. 325–329, 2021.

[23] H. Al Islami, T. Informatika, J. Raya, P. No, and T. Selatan, “Application of the Datamining Association Method Market Basket Analysis Using the Apriori Algorithm to Perform Pattern Analysis of Consumer Spending,” *J. Artif. Intellegence Innov. Appl.*, vol. 1, no. 1, pp. 23–31, 2020, [Online]. Available: http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/69.

[24] J. P. van Wouwe, “A novel association rule mining approach using TID intermediate itemset Enhanced Reader.pdf.” .

[25] M. Rajagukguk, “Implementation of Association Rule Mining to Determine Food Combination Patterns with the Apriori Algorithm,” *J. Fasilkom*, vol. 10, no. 3, pp. 248–254, 2020, doi: 10.37859/jf.v10i3.2308.