Analysis of Transaction Data for Modeling the Pattern of Goods Purchase Supporting Goods Location

Linda Rosliadewi [1,*], Yusmedi Nurfaizal [2], Retno Waluyo [3], Mohammad Imron [4]

[1,3,4] Program Studi Sistem Informasi, Universitas Amikom Purwokerto, Indonesia
[2] Program Studi Bisnis Digital, Universitas Amikom Purwokerto, Indonesia
[1] lindarosliadewi@gmail.com *; [2] faizal@amikompurwokerto.ac.id; [3] waluyo@amikompurwokerto.ac.id; [4] imron@amikompurwokerto.ac.id
* corresponding author

(Received: April 28, 2020; Revised: July 3, 2020; Accepted: August 16, 2020; Available online: December 1, 2020)

Abstract

Arlinda shop is a shop that sells daily necessities located in Salem, Brebes. Each day, this shop generates more and more data that is not used. The store layout which does not get enough attention will affect the level of sales. This study aimed to process the unused transaction data to obtain purchase patterns, some of the most frequently used algorithms were the apriori algorithm and FP-Growth algorithm to find relationship patterns, however, there was a technical constraint in the recommendation technique used which was frequently ignoring a large collection of items. To overcome this problem, the clustering process was carried out using the K-Medoids algorithm so that the association process became smaller. The test was carried out using RapidMiner with a minimum support of 10% - 30% and a minimum confidence of 70% and the results of recommendations for the layout of the goods with the highest lift ratio, namely if someone buys Nuvo BW then he buys pepsodent act, if someone buys wrapping papers then he buys mamy poko, and if someone buys cereal milo then he buys chitato.

Keywords: Association rule; Clustering; Purchasing Patterns

1. Introduction

The number of supermarkets that continues to grow makes it difficult for supermarket managers to implement better marketing strategies, for that managers must pay attention to the buying patterns made by consumers [1]. Examples of strategies for placing goods on shelves that are tailored to the pattern of purchasing goods to find a strategy that consumers want, especially in providing convenience in choosing goods [2].

From this consumer spending pattern, it can be predicted to make more effective strategic decisions [3]. Data transactions that are fast and can accommodate very large amounts result in the accumulation of large amounts of data, data sets that have accumulated are often not used and will multiply. However, the pile of data can still be used by extracting information from the data set into important information and can be used in making decisions to determine sales strategies. The information generated is of course the result of analysis of sales data processing, existing sales data will be processed or analyzed to determine the level of consumer tendencies in each purchase. From the data processing, a pattern of public consumption will be obtained [4].

The Arlinda store has transaction data on sales of goods which continues to grow every day and the transaction data is only stored without knowing the benefits of the data for other things. The collected data pile can produce a buying pattern so that from this pattern it can be known customer interests and shopping habits. Patterns that can be used to find relationships between items so that they can be used to find items whose placement must be close [5]. Judging from the layout of the store and the results of interviews so far, the arrangement of the layout of the goods does not yet have a strategy for the layout of goods based on the goods that are often purchased and the behavior of consumers in buying goods simultaneously at one time. This will certainly affect the level of sales of goods [6].

Based on research from Masnur [7], there are 3 rules of goods formed, namely Milo Active with ABC Kopi Susu, Dancow 1+honey with Ice Cream Corneto, Siip Roasted with Davos Strong then compared with transaction data at Arlinda's shop which has similar goods. There are 4 transactions that have the appropriate goods relationship, namely Milo Active with ABC Kopi Susu, 2 transactions in January and 2 transactions in February. There were 28 transactions in January and February 39 transactions that were not in accordance with the related goods in the previous study. There are several linkages of goods that are not in accordance with the relationship of goods in the Arlinda store, it is necessary to do research on the layout of goods in accordance with consumer spending patterns. So that business actors can determine the right strategy in increasing product sales.

Association rule mining is a data mining method that focuses on finding purchasing patterns by extracting associations or events from transactional data from a store [8]. There are a number of algorithms to solve the Association rule, while in this study the algorithm used is FP Growth which is included in the yahoo association in data mining. The FP Growth algorithm is a development of the a priori algorithm, where there are improvements or developments made so that the FP Growth algorithm has advantages in terms of accuracy and speed in processing data (Rusdiaman, et al, 2018). In the association rules there are technical problems related to the most common recommendation technique, namely the rules about large items. The results of product recommendations are not accurate because retail product information is not available [9].

To overcome this, a clustering process is carried out using the K-Medoid algorithm which is not affected by data outliers so that the clustering results become more accurate [10]. Based on trials that have been carried out by Sumangkut et al. [11] with the title Analysis of Shopping Patterns Using the FP-Growth Algorithm, Self Organizing Map (SOM) and K-Medoids which analyze shopping patterns of shopping patterns using the Fold-Growth algorithm, Self Organizing Map and K Medoid. The research involves supermarket transaction data which has been clustered first, which is 5 clusters. Then the next process is associated with measuring support, confidence and lift ratio. The clustering process is carried out so that the associated data becomes smaller so that the pattern obtained becomes more accurate. In this study, information mining will be carried out using sales transaction data at the Arlinda store by applying the K-Medoids Algorithm for the clustering process and FP-Growth for the association process.

2. Literature Review

2.1. Data Mining

Data mining is defined as the process of finding patterns in data. This process is automated or often semi-automatic. The pattern found must be meaningful. The data needed is usually in large quantities [12]. Data mining is divided into several groups based on the tasks that can be divided, namely [13].

Description Description is a pattern and trend often provides possible explanations for a pattern and trend. Estimation, almost the same as classification, except that the target variable estimates are more numerical than categorical. The model is built using a complete record that provides the value of the predictive variable. Prediction, Prediction is almost the same as classification and estimation, except that in predicting the value of the outcome will be in the future. Classification, In classification, there is a categorical target variable. For example, the classification of income can be separated into three categories, namely income can be separated into three categories, namely high income, medium income, low income. Clustering, namely grouping records, observing or observing and forming classes of objects that have similarities with one another and have similarities with one another and have dissimilarities with records in other clusters. Association, The task of association in data mining is to find attributes that appear at one time. In the business world it is more commonly called shopping cart analysis.

2.1.1. Data Selection

Selection of data from a set of operational data needs to be done before the stage of extracting information in KDD begins. The selected data that will be used for the data mining process, we prefer what kind of data we need for further processing and then the data is stored in a separate file from the operational database so as to provide convenience for subsequent use.

2.1.2. Preprocessing

In general, the data obtained, both from a company's database and experiments, have imperfect entries such as missing data, invalid data or just a typo. In addition, there are also irrelevant data attributes that are also better removed because their presence can reduce the quality or accuracy of the data mining results later. "Garbage in garbage out" (only waste that will be generated if the input is also garbage) is a term that is often used to describe this stage. Data cleaning also affects the performance of the data mining system because the data handled will be reduced in number and complexity.

2.1.3. Transformation

Some data mining techniques require special data formats before they can be applied. For example, some standard techniques such as association analysis and clustering can only accept categorical inputs. Therefore, data in the form of continuous numeric numbers needs to be divided into several intervals. This process is often called binning. Here is also done the selection of data required by the data mining techniques used. This transformation and selection of data also discovers the quality of certain data mining techniques that depend on this stage.

2.1.4. Mining

Data mining is the process of looking for interesting patterns or information in selected data using certain techniques or methods. Techniques, methods, or algorithms in data mining vary widely. The selection of the right method or algorithm really depends on the objectives and the overall KDD process. The data used to become a good model should ideally be sufficient as research data. The more data and the fewer errors, the better the model that is used as a benchmark.

2.1.5. Evaluation

The pattern of information generated from the data mining process needs to be displayed in a form that is easily understood by interested parties. This stage is part of the KDD process called interpretation. This stage includes checking whether the patterns or information found contradict the facts or pre-existing hypotheses.

2.2. Layout

Layout is one of the decisions regarding the layout of production facilities for the efficiency of the company's operations in the long term. The layout has strategic implications as it creates competitive priorities with respect to capacity, process, flexibility and cost. An effective layout can help organizations achieve strategies that support differentiation, low costs, and responsiveness [14].

2.3. Clustering

According to Hair [15] Clustering (Clustering) is one of the tools in data mining that aims to group objects into clusters. By using clustering, you can identify dense areas, find overall distribution patterns, and find interesting relationships between data attributes. According to Hair [15], there are several important types of clustering, namely:

2.3.1. Partial vs Hierarchical

Partial clustering is the division of data objects into non-overlapping subsets (clusters) so that each data object is in exactly one subset. Hierarchical clustering is a set of nested clusters arranged as a hierarchical tree. Each node (cluster) in the tree (except the leaf node) is a combination of its children (cluster) and the root node contains all objects.

2.3.2. Exclusive vs non-Exclusive

Exclusive clustering where each object is in exactly one cluster. On the other hand, in overlapping or non-exclusive clustering, an object can be in more than one cluster simultaneously.

2.3.3. Fuzzy vs non-Fuzzy

In fuzzy clustering, a point is included in each cluster with a weight value between 0 and 1. The sum of these weights is equal to 1. The probability clusters have the same characteristics.

2.3.4. Partial vs Complete

In complete clustering, each object is placed in a cluster but in partial clustering, not all objects are placed in a cluster. There may be inappropriate objects to be placed in one of the clusters, for example in the form of outliers or noise.

2.4. Association Analysis

Association analysis or association rule mining is a data mining technique to find association rules between a combination of items [16].

The importance of an association rule can be determined by two parameters, support (supporting value), namely the percentage of the combination of these items in the database and confidence (certainty value), which is the strong relationship between items in the association rules [17]. Association rules are usually stated in the form [18]:

$$\{Bread,\ Butter\} \rightarrow \{Milk\}(Support\ =\ 40\%,\ Confidence\ =\ 50\%) \qquad (1)$$

Which means: “50% of the transactions in the database containing bread and butter items also contain dairy items. Meanwhile, 40% of all transactions in the database contain those three items”. It can also be interpreted: “A consumer who buys bread and butter has a 50% chance of also buying milk. This rule is quite significant because it represents 40% of the transaction records so far”. Association analysis is defined as a process to find all association rules that meet the minimum requirements for support (minimum support) and minimum requirements for confidence (minimum confidence). The basic method of analysis is divided into two stages [19]:

2.4.1. High Frequency Pattern Analysis

This stage is looking for a combination of items that meet the minimum requirements of the support value in the database. The support value of an item is obtained by the following formula:

ISSN 2723-6471

$$support(A) = \frac{The\ transaction\ amount\ contains\ A}{Total\ transactions} \quad (2)$$

While the support value of the 2 items is obtained from the following formula:

$$support(A \cap B) = \frac{The\ number\ of\ transactions\ contains\ A\ and\ B}{Total\ transactions} \quad (3)$$

2.4.2. Associative Rule Formation

After all high-frequency patterns are found, then look for the associative rule that meets the minimum requirements for confidence by calculating the confidence of the associative rule A → B. The confidence value of the rule A → B is obtained from the following formula:

$$confidence\ = P(A) = \frac{The\ number\ of\ transactions\ contains\ A\ and\ B}{Total\ transactions\ contains\ A} \quad (4)$$

The FP Growth Algorithm is one of the alternative algorithms that can be used to determine the data sets that most often appear simultaneously, extracting frequently occurring itemsets (frequent) will be done by generating a Tree or FP-Tree data structure. The characteristic of the FP-Growth algorithm is that the data structure used is called FP-Tree, so the tree data structure generation will be carried out on frequent itemset excavation.

3. Research Model

The design of the research flow can be seen as shown in Figure 1.

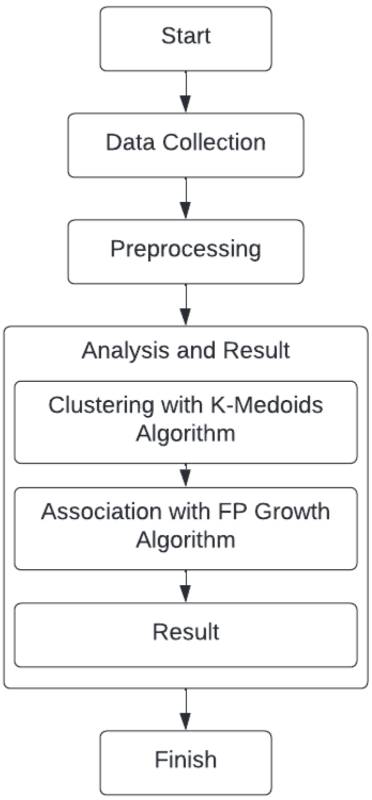

Figure. 1. Research Flow

3.1. Data Collection

The data taken is sales transaction data at the Arlinda store. The form of data received is in the form of an excel file. Sales data will be carried out in the data cleaning stage or deletion of unused attributes to match what is needed by the system.

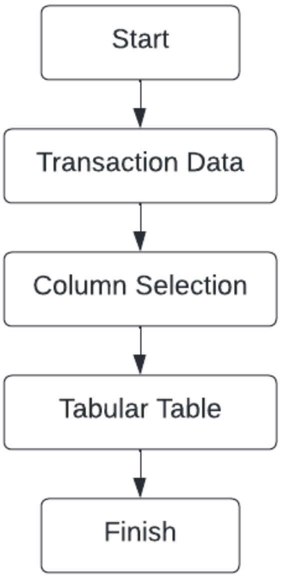

Figure. 2. Preprocessing Stage

3.2. Preprocessing Stage

At this stage the sales data recap will be carried out in the stage of cleaning data or deleting unused attributes and adjusting the data format so that it can be processed with RapidMiner, namely by converting the data into tabular form. Based on Figure 2, the first steps to be carried out at the preprocessing stage are as follows:

3.2.1. Attribute Column Selection

Arlinda store transaction data for January and February 2019 consisted of 11 different attributes. In the analysis process, the author does not need all the existing attributes, only uses the Item Name table to be processed in RapidMiner.

3.2.2. Tabular Table

Changes in data are converted into appropriate datasets so that they can be used in the association process, namely making the dataset numeric, 1 indicates that an item is in the buyer's shopping cart and 0 indicates that an item is not in the buyer's shopping basket.

3.3. Stages of Analysis and Results

In this study, using the K-Medoids algorithm and the FP-Growth algorithm. The stages in the application of the algorithm carried out include:

3.3.1. K-Medoid Clustering Algorithm

Starting from the data clustering process by entering the number of data and the number of clusters that have been determined. The clustering process aims to make the associated data smaller so that the resulting pattern can be more accurate.

3.3.2. Association with FP-Growth Algorithm

Each of the clustering stages with the K-Medoids algorithm that is formed will then carry out an association process. The association process is carried out using the FP-Growth algorithm to generate rules and is measured using support and confidence.

3.4. Results

The results of the research are in the form of rules obtained after the data mining process is completed. Then conclude the resulting rule and can provide recommendations for the placement of goods.

4. Result and Discussion

4.1. Data

The dataset uses transaction data from the Arlinda Store taken from January to February 2019. There are 448 products and 2803 transactions that will be used in this study. The following is an example of a sales recap at the Arlinda store in the form of an excel file with .xlsx format.

	A	B	C	D	E	F	G	H	I	J	K
1	No	Tanggal	No. Transaksi	Kode	Nama	hBeli	Harga	Jml	TotBeli	Totharga	Laba
2	1	01/01/2019	S-ARL/0119/1	500019	CUSSONS BABY CRM SOFT&SM	13,600	16,000	1	13,600	16,000	2,400
3	2	01/01/2019	S-ARL/0119/2	400085	PROMINA KACANG HIJAU 120G	11,008	12,500	1	11,008	12,500	1,492
4	3	01/01/2019	S-ARL/0119/3	402613	CHOCOLATOS DARK MINI 33G	1,700	2,500	1	1,700	2,500	800
5	4	01/01/2019	S-ARL/0119/3	300366	SO KLIN LANTAI 400ML KNG	4,075	6,500	1	4,075	6,500	2,425
6	5	01/01/2019	S-ARL/0119/4	500054	CUSSONS COTTON BUDS 100'	3,685	4,500	1	3,685	4,500	815
7	6	01/01/2019	S-ARL/0119/5	100369	CUSSONS KID HAIR&BODY COL	9,020	11,000	1	9,020	11,000	1,980
8	7	01/01/2019	S-ARL/0119/5	402442	POTA BEE 15G RUMPUT LAUT	1,575	2,000	2	3,150	4,000	850
9	8	01/01/2019	S-ARL/0119/5	600152	TELON MY BABY+ 90ML	20,515	25,000	1	20,515	25,000	4,485
10	9	01/01/2019	S-ARL/0119/6	401129	CORNETTO MINI	2,500	3,000	1	2,500	3,000	500
11	10	01/01/2019	S-ARL/0119/7	202998	SUNSILK 10ML BLACK	4,650	5,500	1	4,650	5,500	850
12	11	01/01/2019	S-ARL/0119/8	100909	MY BABY TELON PLUS 150ML	32,615	39,000	1	32,615	39,000	6,385
13	12	01/01/2019	S-ARL/0119/8	100035	MY BABY PWD SWEET FLORAL	2,700	3,500	1	2,700	3,500	800
14	13	01/01/2019	S-ARL/0119/8	405034	KAPAL API SPC MIX 25GR	996	1,100	10	9,960	11,000	1,040
15	14	01/01/2019	S-ARL/0119/9	200947	MARINA HBL 200ML UW FRESH \	7,460	9,000	1	7,460	9,000	1,540
16	15	01/01/2019	S-ARL/0119/9	201569	SO KLIN PW 900ML REF UNGU	8,500	10,000	1	8,500	10,000	1,500
17	16	01/01/2019	S-ARL/0119/9	101127	VEET 90ML NORMAL	11,950	13,000	1	11,950	13,000	1,050
18	17	01/01/2019	S-ARL/0119/9	202800	CIPTADENT SG CC	2,350	3,000	1	2,350	3,000	650
19	18	01/01/2019	S-ARL/0119/10	600152	TELON MY BABY+ 90ML	20,515	25,000	1	20,515	25,000	4,485
20	19	01/01/2019	S-ARL/0119/10	502690	SWEETY SILVER PANTS M30	54,000	63,000	1	54,000	63,000	9,000
21	20	01/01/2019	S-ARL/0119/10	500876	SWEETY SILVER PANTS M18+2	35,500	40,000	1	35,500	40,000	4,500
22	21	01/01/2019	S-ARL/0119/11	203359	SERASOFT SHAMPO HAIRFALLf	4,620	5,500	1	4,620	5,500	880
23	22	01/01/2019	S-ARL/0119/11	500264	MITU BABY REGULAR 24'S CHAN	5,500	7,500	1	5,500	7,500	2,000
24	23	01/01/2019	S-ARL/0119/11	300109	BAGUS KPR AJAIB 7GR JUMBO	2,750	3,500	1	2,750	3,500	750

Figure. 3. Transaction Data

4.2. Dataset

The raw data that will be read by the system is made into a tabular table first, then the dataset reading is generated in the image below.

ISSN 2723-6471

No	FEAST CHOCO LATO	FEMI NINE COM FORT 10PDS	FRES & NATU RAL SUMMER 100ML	FRESH CARE GREEN TEA	FRISIAN FLAG COK SCHT	...	KO DO MO PG 4GR
31	0	0	1	0	0	...	0
32	0	0	0	0	0	...	0
33	0	0	0	0	0	...	0
34	0	0	0	0	0	...	0
35	0	0	1	0	1	...	1
36	0	0	0	0	0	...	0
37	0	0	0	0	0	...	0
38	0	0	0	0	0	...	1
39	1	0	0	0	0	...	0
...	...	...	...	...	...	...	...
2278	0	1	0	0	0	...	0
2279	0	0	0	0	0	...	0
2280	0	0	0	0	0	...	0
2281	0	0	0	0	0	...	0
2282	0	0	1	0	0	...	0

Figure. 4. Tabular Table

4.3. K-Medoids Process

After preprocessing by creating a tabular table, the dataset is then processed using the K-Medoids algorithm. In this study, the number of clusters determined was 5 clusters using the RapidMiner application.

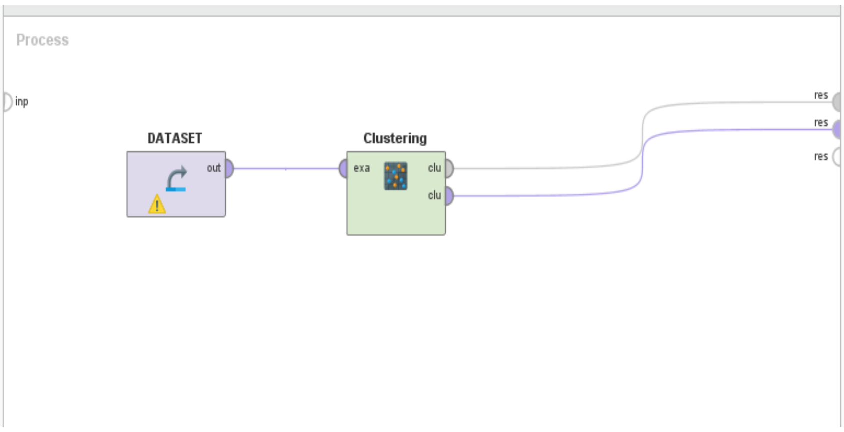

Figure. 5. Clustering process

The dataset in excel form that has been imported into RapidMiner is connected to the clustering operator and divides the data into 5 clusters (k). The results of the K-Medoids process can be seen in Figure 6.

Row No.	No	cluster	AINIE BRS J...	AINIE BRS J...	AMANDA MA...	AMANDA MA...	ANLENE ACT...	ANLENE ACT...	ARA
1	1	cluster_0	1	0	1	0	1	0	1
2	2	cluster_0	1	0	1	0	1	0	1
3	3	cluster_0	1	0	1	0	1	0	1
4	4	cluster_0	1	0	1	0	1	0	1
5	5	cluster_0	1	0	1	0	1	0	1
6	6	cluster_0	1	0	1	0	1	0	1
7	7	cluster_0	1	0	1	0	1	0	1
8	8	cluster_0	1	0	1	0	1	0	1
9	9	cluster_0	1	0	1	0	1	0	1
10	10	cluster_0	1	0	1	0	1	0	1
11	11	cluster_0	1	0	1	0	1	0	1
12	12	cluster_0	1	0	1	0	1	0	1
13	13	cluster_0	1	0	1	0	1	0	1
14	14	cluster_0	1	0	1	0	1	0	1

ExampleSet (2,803 examples, 2 special attributes, 878 regular attributes)

Figure. 6. Clustering process results

The clustering results obtained are cluster 0 as many as 506 items, cluster 1 as many as 1326, cluster 2 as many as 858, cluster 3 as many as 85 and cluster 4 as many as 26 items then the association process will be carried out using FP-Growth.

4.4. FP-Growth Process

After each of the clusters is formed, then do the association using the FP-Growth algorithm. The results of this association process are measured by the value of support and confidence. In this calculation, experiments were carried out with a minimum support of 10%-30% and a minimum confidence of 70% using the RapidMiner application.

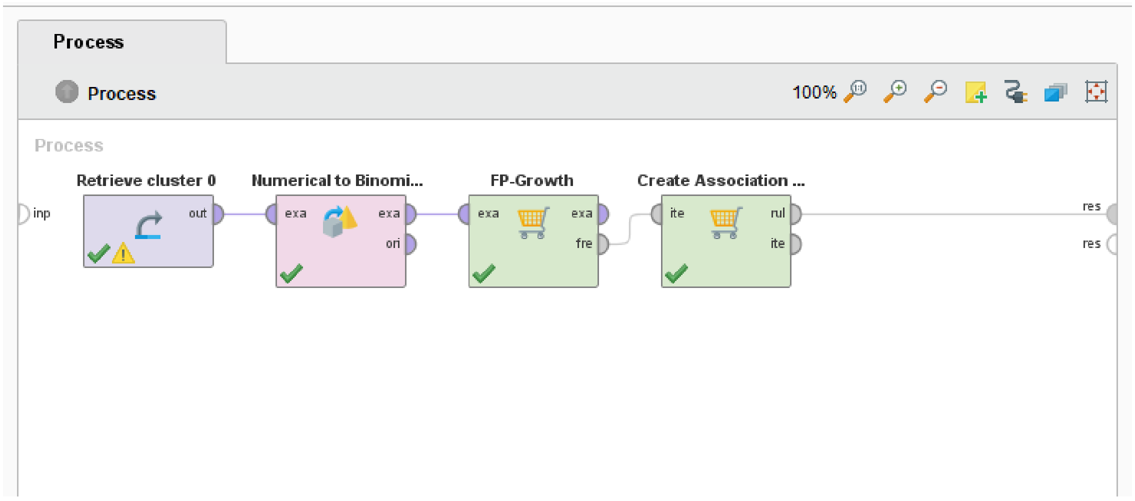

Figure. 6. Association process

In this FP-Growth process, cluster 1 to cluster 5 is associated with each cluster. The results of the association of cluster 0, cluster 1 and cluster 2 with a minimum support of 10% to 30% and a minimum confidence of 70% are the same as clusters 0 and 1, namely that no rules or patterns are found. Results From the search for rules in cluster 3 data, there are 3 rules or patterns that are formed, then the search for rules in cluster 4 is found as much as 1 rule with a minimum support of 20-30%. The results that are formed are based on clustering 1 to 5 clusters where there are rules with the highest support and confidence, namely in cluster 3 which states that there are 4 rules and cluster 4 which has 1 rule.

Then the results of these 4 rules are taken based on the support value, confidence value and lift ratio which is more than 1 can be seen in table 1. According to Gepp et al. [20] the lift ratio value is one way to show whether a rule is valid or not. . Then the results of this calculation can be used as recommendations to be used as rules in placing goods.

Table. 1. Results formed

Itemset	*Sup*	*Conf*	*Lift ratio*
If you buy soap then buy detergent	0,035	0,750	5,795
If you buy soap then buy toothpaste	0,035	0,750	7,969
If you buy baby oil then buy diapers	0,035	0,750	9,107
If you buy cereal then buy milk	0,035	0,750	1,566

5. Conclusion

The calculation using the K-Medoids algorithm can be applied to a large number of datasets so that it can assist in the process of searching for association rules in order to find more accurate product recommendations because the associated dataset becomes smaller. After the calculation process is carried out with a minimum support of 10% to 30% association rules that are formed only in cluster 3 and cluster 4. There are 4 rules that have the highest value, namely if you buy shampoo then buy detergent, if you buy soap then buy toothpaste, if you buy baby oil then buy diapers, if you buy cereal then buy milk. The pattern that has been formed can be used by the store to help make decisions in determining the layout of the goods. Further research can add data with different conditions, for example data in the rainy season and dry season or in the highlands and in the lowlands. Experiment with clustering algorithms and other associations.

References

[1] G. Gunadi and I. S. Dana, “Application of Data Mining Market Basket Analysis Method on Sales Data for Book Products Using Apriori Algorithm and Frequent Pattern Growth (FP-Growth): A Case Study of PT,” Gramedia". J. Telemat. MKOM, vol. 4, no. 1.

[2] H. Aprizal and N. Wahyu, “Implementation of Data Mining for Positioning Goods on Shelves Using the Apriori Method at PT Midi Utama Indonesia",” Techo.com, vol. 15, p. 4.

[3] K. Christidis, D. Apostolou, and G. Mentzas, “Exploring Customer Preferences With Probabilistic Topics Models".”

[4] S. Defiyanti, M. Jajuli, and R. W. Nurul, “Optimizing K-Medoids in Clustering Scholarship Applicants Using Cubic Clustering Criterion”.” .

[5] M. I. Ghozali, Z. E. Ri’fan, and H. S. Wibowo, “Analysis of Shopping Patterns Using the FP-Growth Algorithm, Self Organizing Map (SOM) and K Medoids",” J. SIMETRIS Vol.8 No.1.

[6] N. Maharani et al., “Implementation of Data Mining for Mini Market Layout Settings by Applying Association rules”,” J. Comput.

[7] A. Masnur, Data Mining Analysis Using Market Basket Analysis to Determine Consumer Buying Patterns". SATIN- Science and Information Technology Vol.1 No.2.

[8] D. Nofriansyah, The Concept of Data Mining VS Decision Support Systems. Yogyakarta: Depublish.

[9] L. I. Prahartiwi, “Frequent Itemset Search in Shopping Cart Analysis Using the FP-Growth Algorithm”,” Inf. Syst. Educ. Prof., vol. 2, no. 1.

[10] H. Santoso, I. Putu, and P. H., “Data Mining Analysis of Product Purchase Patterns Using the Apriori Algorithm Method",” in 2016 National Seminar on Information Technology and Multimedia, .

[11] K. Sumangkut, L. Ari, and T. Virginia, “Analysis of the Daily Mart supermarket shopping pattern to determine the layout of goods using the FP-Growth Algorithm",” E-Journal Informatics Eng., vol. 8.

[12] K. N. Wijaya, Shopping Cart Frequency Pattern Analysis Using Apriori Algorithm. Case Study: Minimarket.

[13] N. Omar, Z. ‘Amirah Johari, and M. Smith, “Predicting fraudulent financial reporting using artificial neural network,” J. Financ. Crime, vol. 24, no. 2, pp. 362–387, Jan. 2017, doi: 10.1108/JFC-11-2015-0061.

[14] U. C. Moharana, S. P. Sarmah, and P. K. Rathore, “Application of data mining for spare parts information in maintenance schedule: a case study,” J. Manuf. Technol. Manag., vol. 30, no. 7, pp. 1055–1072, Jan. 2019, doi: 10.1108/JMTM-09-2018-0303.

[15] J. F. Hair, “Knowledge creation in marketing: the role of predictive analytics,” Eur. Bus. Rev., vol. 19, no. 4, pp. 303–315, Jan. 2007, doi: 10.1108/09555340710760134.

[16] G. Boskou, E. Kirkos, and C. Spathis, “Classifying internal audit quality using textual analysis: the case of auditor selection,” Manag. Audit. J., vol. 34, no. 8, pp. 924–950, Jan. 2019, doi: 10.1108/MAJ-01-2018-1785.

[17] M. Yu and C. Guo, “Using news to predict Chinese medicinal material price index movements,” Ind. Manag. Data Syst., vol. 118, no. 5, pp. 998–1017, Jan. 2018, doi: 10.1108/IMDS-06-2017-0287.

[18] J. Manickavasagam and V. S., “An investigational analysis on forecasting intraday values,” Benchmarking An Int. J., vol. 27, no. 2, pp. 592–605, Jan. 2020, doi: 10.1108/BIJ-11-2018-0361.

[19] X. Ji, J. Wang, and Z. Yan, “A stock price prediction method based on deep learning technology,” Int. J. Crowd Sci., vol. 5, no. 1, pp. 55–72, Jan. 2021, doi: 10.1108/IJCS-05-2020-0012.

[20] A. Gepp, M. K. Linnenluecke, T. J. O’Neill, and T. Smith, “Big data techniques in auditing research and practice: Current trends and future opportunities,” J. Account. Lit., vol. 40, no. 1, pp. 102–115, Jan. 2018, doi: 10.1016/j.acclit.2017.05.003.