Analysis of Apriori Algorithm on Sales Transactions to Arrange Placement of Goods on Minimarket

Putu Bagus Indra Sukadiana Putra[1*], Ni Putu Sri Merta Suryani[2], and Sri Aryani[3]

[1,2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3]Department of Electrical and Computer Engineering, Udayana University
*Email: indra.sukadianaputra@gmail.com

Abstrak — **In the business world, business development is a very important factor to be reckoned with. This is done in order to always survive in the competitive world of business. Improving product quality, adding product types and reducing the company's operating costs are some things that can be done in order to survive in the competition. The thing that can be done to meet those needs is to perform sales transaction data analysis using Data Mining. One of the Algorithms that belongs to an association in Data Mining that matches is the Apriori Algorithm. A priori algorithm aims to find frequent item sets in a set of data. A priori algorithm defined a process to find a priori rules that meet minimum requirements for support and minimum requirements for confidence. This research seeks to develop a business strategy of placement of groceries that is adjusted to the consumer consumption pattern in minimarket. Thus, the arrangement of merchandise can be adjusted with the rules of association to fit the pattern of consumer consumption.**

Keywords : Data Mining, Algoritma Apriori, Minimarket

I. INTRODUCTION

The development and business competition in world trade through free market economy and information technology advancement bring the company to a level of increasingly fierce competition and be more open in meeting the demands of customers who are also higher. Companies must be able to implement good business strategy to compete and still have market share. Competition in business can not be separated from the information technology is a hot topic for now.

Minimarkets are supermarkets that are store size and number of items sold less than supermarkets and hypermarkets. Where the item amount of goods sold for minimarket between 3000 - 5000 kinds of goods. Store size is between 300 -500 m2. In addition the radius area served minimarket also not far that is effectively less than 2 km. Minimarket has spread all over Indonesia. Thus, minimarket entrepreneurs should continue to develop in order to always survive in business competition. Improving product quality, adding product types and reducing the company's operating costs are some things that can be done in order to survive in the competition.

Sales transaction data is one of the things that can be utilized for a business decision making. Most sales transaction data is not re-used, and only stored as archives and only used for making a sales report. Data mining is one of the sciences that can be applied in this case. Sales transactions that are not utilized properly can be extracted and reprocessed into a useful information using data mining techniques. Using one method of data mining, namely Apriori Algorithm, transaction sales data can be processed again so that it can produce a pattern of consumer purchases. This consumer purchase pattern will help the owner to make a business decision.

This research seeks to develop a business strategy of placement of groceries that is adjusted to the consumer consumption pattern in minimarket. Thus, the arrangement of merchandise can be adjusted with the rules of association to fit the pattern of consumer consumption.

II. METHOD

A. DATA MINING

Data mining is a term used to describe the discovery of knowledge in a database [1]. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases [2].

A common definition of data mining itself is the process of searching for hidden patterns in the form of a previously unknown knowledge of a set of data in which data can be in a database, data werehouse, or information storage medium other. Important related issues in data mining are:

1. Data mining is an automated process to existing data.

2. Data to be processed in the form of very large data.

3. The purpose of data mining is to obtain a relationship or pattern that may provide useful indications [2].

B. DATA MINING GROUP

Data mining is divided into several groups based on tasks that can be done [3]:

1. Description

Sometimes researchers and analysts simply want to try to figure out how to describe the patterns and trends contained in the data. For example, the polling officer may not find the information or the fact that who is not professional enough will be little supported in the presidential election. Descriptions of patterns and trends often provide possible explanations for a pattern or trend.

2. Estimates

The estimates are almost identical to the classification, except that the estimation target variable is more nuPepperl than in the direction of the category. The model built with the complete record provides the value of the target variable as the predicted value. Next, in the next review the estimated value of the target variable is based on the predicted variable value.

3. Prediction

The prediction is almost the same as the classification and estimation, except that in prediction the value of the result will be in the future. Several methods and techniques used in the classification and estimation can also be used (for the right circumstances) for prediction.

4. Classification

In the classification, there are target categories of variables. For example, the classification of income can be separated into three categories, namely high income, medium income, and low income.

5. Clustering

Clustering is a grouping of records, observations, or the attention and form of classes of objects that have similarities. The cluster is a collection of records that have similarities with each other and have a non-incompatibility with records in other clusters. The clustering is different from the classification of the absence of the target variable in the clustering. Clustering does not attempt to classify, estimate, or predict the value of the target variable. However, the clustering algorithm tries to divide the entire data into groups that have a similarity record in one group will be maximal, while the similarity with records in other groups will be of minimal value.

6. Association

The task of association in data mining is to find attributes that appear at a time. In the business world more commonly called the analysis of shopping cart (market basket analysis).

C. APRIORI ALGORITHM

Apriori algorithm includes the type of association rules on data mining. In addition to Apriori, which is included in this algorithm is the method of Generalized Rule Induction and Hash Based algorithm. The rules that state the association between several attributes are often called affinity analisys or market basket analysis [2].

The way this algorithm works is that the algorithm will generate new candidates from the k-itemset from the frequent itemset in the previous step and calculate the k-itemset support value. An item that has a support value below of minsup will be deleted. The algorithm ceases when no new generated item frequencies are generated.

From the results of frequent itemset, the next step is calculated minconf follow the formula according to which has been determined. Support does not need to be seen again, because the generated frequent itemset is obtained from viewing its minsup. If the rule that meets the specified limits and restrictions is high, then the rule is classified as strong rules.

The Apriori algorithm is processed iteratively, first recognizing frequentitemset with one item. In each subsequent iteration, the frequentitemset s recognized in the previous iteration is developed with other items to generate larger candidate itemsets. Considering only the itemset obtained through the expansion of the frequentitemset, we can reduce the number of candidate frequentitems; This optimization is essential for efficient execution. Priority assures that this optimization is correct, which means we do not lose frequentitemset. The single search of all transactions is

sufficient to determine the candidate of the itemset generated in one iteration is the frequentitemset s. the algorithm ends if no known frequentitemset is in one iteration.

The Apriori Algorithm Principle is [4]:
1. Collect the number of single items, get the big items.
2. Get the candidate pairs, count => large pairs of items.
3. Get the candidate triplets, count => large triplets of itemitem and so on.
4. As a hint: every subnet of a frequent itemset should be frequent..

According to Kusrini and Luthfi [2], the basic methodology of association analysis is divided into two stages:
1. Analyze the high frequency pattern

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

$$\text{Support} \, (\, A \,) \, = \, \frac{\textbf{Amount of transactions contains A}}{\textbf{Amount of transactions}}$$

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

$$\text{Support} \, (\, A \cap B \,) \, = \, \frac{\textbf{Amount of transactions contains A and B}}{\textbf{Amount of transactions}}$$

2. The establishment of associative rules

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

$$\text{Confidence} = P \, (B|A) \, = \, \frac{\textbf{Amount of transactions contains A and B}}{\textbf{Amount of transactions contains A}}$$

III. RESULT AND DISCUSSION

A. DATA COLLECTION

In this study the source data used comes from the sales transaction data on a minimarket. The data will be processed to generate knowledge that can be used as a business strategy development. Initial data obtained from the minimarket is still a random sale transaction that has not been selected into ready-to-use data.

B. DATA SELECTION

After the data is collected then the selection of sales transaction data has been obtained from the minimarket. At this stage do the sorting of sales transaction data because not all the data will be used in this research. Sales transaction data to be

taken depends on the amount of goods or items purchased by consumers. The amount of goods in sales transaction data to be taken is the transaction data with the amount of goods two and three goods only.

C. PRE-PROCESSING / CLEANING

Once selected, the sales transaction data is done preprocessing / cleaning is the process of cleaning attributes that are not used in the process of data mining and only a few attributes that can be used. In sales transaction data there are several attributes such as transaction number, transaction date, item code, item name, quantity of goods, price of goods, price of goods subtotal, total purchase price. After cleaning process the remaining attributes are transaction number and item name. In the name of the goods attribute is rounding the name of the goods because in the goods which is a type of goods but there are differences such as the size, brand, and so it is rounded the name to make it easier in the data mining process analysis will be done.

D. TRANSFORMASI DATA

At this stage data transformation in accordance with the system to be used in data mining analysis.

TABEL I. List of Sales Transactions

No.	Transaction Number	Name Of Goods
1	20181008	Bread, Jam
2	20181010	Chocolate Powder , Salt
3	20181011	ABC Chili Sauce, ABC Tomato Sauce, Soy Sauce
4	20181012	Condensed Milk, Biscuits
5	20181019	Soy Sauce, ABC Tomato Sauce
6	20181020	Pepper, ABC Tomato Sauce, ABC Chili Sauce
7	20181021	ABC Chili Sauce, Salty Soy Sauce, Flavoring
8	20181022	Rice Flour, Wheat Flour
9	20181024	Sugar, ABC Chili Sauce, ABC Tomato Sauce
10	20181025	Meses, Body lotion
11	20181026	Seasoning Sachet, Vinegar
12	20181027	Vanili, Vinegar, ABC Tomato Sauce
13	20181028	Mayonaise, ABC Chili Sauce, ABC Tomato Sauce
14	20181029	Sugar, ABC Tomato Sauce
15	20181030	ABC Chili Sauce, ABC Tomato Sauce, Soy Sauce
16	20181032	Salty Soy Sauce, Soy Sauce, ABC Tomato Sauce
17	20181034	Condensed Milk, Meses
18	20181036	Bread, Jam, Cheese
19	20181037	Food Coloring, Vanili, Cooking Oil
20	20181038	Flavoring, ABC Chili Sauce
21	20181039	Shampoo, Liquid Bath Soap, ABC Chili Sauce
22	20181042	Pomade, Pepper, ABC Tomato Sauce
23	20181043	Cooking Oil, Soy Sauce, ABC Tomato Sauce
24	20181044	Biscuits, Condensed Milk, Meses
25	20181045	Chocolate Powder , Meses

No.	Transaction Number	Name Of Goods
26	20181049	Bread Crumb, Wheat Flour, Deodoran
27	20181055	Rice Flour, Mayonaise, Wheat Flour
28	20181056	Mayonaise, Salty Soy Sauce, ABC Chili Sauce
29	20181057	Jam, Bread, Condensed Milk
30	20181058	Chili Powder, Salt, ABC Chili Sauce
31	20181063	Cooking Oil, Salt, Sugar
32	20181064	Vanili, ABC Tomato Sauce, Sugar
33	20181066	ABC Chili Sauce, ABC Tomato Sauce
34	20181068	Flavoring, Mayonaise
35	20181071	Salty Soy Sauce, ABC Chili Sauce, Salt
36	20181073	Rice Flour, Wheat Flour
37	20181074	Salt, Soy Sauce, ABC Chili Sauce
38	20181076	Salty Soy Sauce, Mayonaise, Vanili
39	20181077	Bread, Jam
40	20181078	Vinegar, Mayonaise, Salty Soy Sauce
41	20181079	Chili Powder, ABC Tomato Sauce
42	20181080	Cooking Oil, Food Coloring
43	20181081	Condensed Milk, Meses, Cheese
44	20181083	Bread Crumb, ABC Tomato Sauce
45	20181085	Wheat Flour, Sugar, ABC Chili Sauce
46	20181086	ABC Tomato Sauce, ABC Chili Sauce, Mayonaise
47	20181088	Liquid Bath Soap, Shampoo
48	20181092	Wheat Flour, ABC Chili Sauce
49	20181095	Jam, Bread
50	20181098	Salt, Soy Sauce

The principle of extracting minimarket transaction data using a priori algorithm is:
1. Determine the transaction limit and specify the list of 1-itemset candidates. The amount of the transaction limit is 3.

TABEL II. List of Candidates 1-itemset

No.	Name Of Goods	Amount
1	Bread	5
2	Jam	5
3	Soy Sauce	7
4	ABC Chili Sauce	16
5	ABC Tomato Sauce	16
6	Mayonaise	7
7	Salty Soy Sauce	6
8	Meses	5
9	Condensed Milk	5
10	Rice Flour	3
11	Wheat Flour	6
12	Salt	6
13	Sugar	5

2. The list of 1-itemset above, made into a frequent 2-itemset list, will get results:

TABEL III. List of 2-itemset Candidates

No.	Name Of Goods	Amount
1	Bread , Jam	5
2	Soy Sauce, ABC Chili Sauce	3
3	Soy Sauce, ABC Tomato Sauce	5
4	Mayonaise, Salty Soy Sauce	3
5	Salty Soy Sauce, mayonaise	3
6	ABC Chili Sauce, mayonaise	2
7	Meses, Condensed Milk	3
8	Rice Flour, Wheat Flour	3
9	Salty Soy Sauce, ABC Chili Sauce	3
10	Salt, ABC Chili Sauce	3
11	ABC Chili Sauce, ABC Tomato Sauce	7
12	Sugar, ABC Tomato Sauce	3

3. Calculate the value of support and confidence of each frequent itemset so that the candidates appear association rules. To calculate the support and confidence used the formula:

$$\text{Support}\,(\,A \cap B\,) = \frac{\text{Amount of transactions contains A and B}}{\text{Amount of transactions}}$$

$$\text{Confidence} = P\,(B|A) = \frac{\text{Amount of transactions contains A and B}}{\text{Amount of transactions contains A}}$$

The calculation obtained is:

1. Support and Confidence for Bread and Jam:

Support BREAD and JAM $= \frac{5}{50} = 0.1$ x 100% = 10%

Confidence BREAD and JAM $= \frac{5}{5} = 1$ x 100% = 100%

2. Support and Confidence for Soy Sauce and ABC Chili Sauce:

Support SOY SAUCE and ABC CHILI SAUCE $= \frac{3}{50} = 0.06$ x 100% = 6%

Confidence SOY SAUCE and ABC CHILI SAUCE $= \frac{3}{7} = 0.42$ x 100% = 42%

3. Support and Confidence for Soy Sauce and ABC Tomato Sauce:

Support SOY SAUCE and ABC TOMATO SAUCE $= \frac{5}{50} = 0.1$ x 100% = 10%

Confidence SOY SAUCE and ABC TOMATO SAUCE $= \frac{5}{7} = 0.71$ x 100% = 71%

4. Support and Confidence for Mayonaise and Salty Soy Sauce:

Support MAYONAISE and SALTY SOY SAUCE $= \frac{3}{50} = 0.06$ x 100% = 6%

Confidence MAYONAISE and SALTY SOY SAUCE $= \frac{3}{7} = 0.42$ x 100% = 42%

5. Support and Confidence for Salty Soy Sauce and Mayonaise:

Support SALTY SOY SAUCE and MAYONAISE $= \frac{3}{50} = 0.06$ x 100% = 6%

Confidence SALTY SOY SAUCE and MAYONAISE $= \frac{3}{6} = 0.5$ x 100% = 50%

6. Support and Confidence for ABC Chili Sauce and Mayonaise:

Support ABC CHILI SAUCE and MAYONAISE $= \frac{2}{50} = 0.04$ x 100% = 4%

Confidence ABC CHILI SAUCE and MAYONAISE $= \frac{2}{16} = 0.12$ x 100% = 12%

7. Support and Confidence for Meses and Condensed Milk:

Support MESES and CONDENSED MILK $= \frac{3}{50} = 0.06$ x 100% = 6%

Confidence MESES and CONDENSED MILK $= \frac{3}{5} = 0.6$ x 100% = 60%

8. Support and Confidence for Rice Flour and Wheat Flour:

Support RICE FLOUR and WHEAT FLOUR $= \frac{3}{50} = 0.06$ x 100% = 6%

Confidence RICE FLOUR and WHEAT FLOUR $= \frac{3}{3} = 1$ x 100% = 100%

9. Support and Confidence for Salty Soy Sauce and ABC Chili Sauce:

Support SALTY SOY SAUCE and ABC CHILI SAUCE $= \frac{3}{50} = 0.06$ x 100% = 6%

Confidence SALTY SOY SAUCE and ABC CHILI SAUCE $= \frac{3}{6} = 0.5$ x 100% = 50%

10. Support and Confidence for Salt and ABC Chili Sauce:

Support SALT and ABC CHILI SAUCE $= \frac{3}{50} = 0.06$ x 100% = 6%

Confidence SALT and ABC CHILI SAUCE $= \frac{3}{6} = 0.5$ x 100% = 50%

11. Support and Confidence for ABC Chili Sauce and ABC Tomato Sauce:

Support ABC CHILI SAUCE and ABC TOMATO SAUCE $= \frac{7}{50} = 0.14$ x 100% = 14%

Confidence ABC CHILI SAUCE and ABC TOMATO SAUCE = $\frac{7}{16}$ = 0.43 x 100% = 43%

12. Support and Confidence for Sugar and ABC Tomato Sauce:

Support SUGAR and ABC TOMATO SAUCE = $\frac{3}{50}$ = 0.06 x 100% = 6%

Confidence SUGAR and ABC TOMATO SAUCE = $\frac{3}{16}$ = 0.19 x 100% = 19%

TABEL IV. Results Process Analysis With Supreme Support Value

No	Rules	Support (%)	Confidence (%)
1	If buying " Bread " then it will buy " Jam "	10	100
2	If buying " Soy Sauce " then it will buy " ABC Chili Sauce "	6	42
3	If buying " Soy Sauce " then it will buy " ABC Tomato Sauce "	10	71
4	If buying " Mayonaise " then it will buy " Salty Soy Sauce "	6	42
5	If buying " Salty Soy Sauce " then it will buy " Mayonaise "	6	50
6	If buying " ABC Chili Sauce " then it will buy " Mayonaise "	4	12
7	If buying " Meses " then it will buy " Condensed Milk "	6	60
8	If buying " Rice Flour " then it will buy " Wheat Flour "	6	100
9	If buying " Salty Soy Sauce " then it will buy " ABC Chili Sauce "	6	50
10	If buying " Salt " then it will buy " ABC Chili Sauce "	6	50
11	If buying " ABC Chili Sauce " then it will buy " ABC Tomato Sauce "	14	43
12	If buying " Sugar " then it will buy " ABC Tomato Sauce "	6	19

No.	Placement of Goods
2	Bread, Jam, Condensed Milk, Biscuits, Cheese, Meses, Chocolate Powder
3	Rice Flour, Wheat Flour, Sugar, Cooking Oil, Vanili, Food Coloring, Bread Crumb
4	Shampoo, Liquid Bath Soap, Deodoran, Pomade, Body lotion
5	Soy Sauce, Salty Soy Sauce, Flavoring, Vinegar, Seasoning Sachet

IV. CONCLUSION

Research can be concluded that the data mining and Apriori algorithm technique can be implemented on sales transaction data, one of them to get the pattern of consumer purchasing. Patterns obtained, will produce a rule or rules about the relevance of a product. The above analysis can be used for various purposes in the sale, one of which is to set the placement of goods or the layout of goods. Layout settings can be done with transactions that have a high confidence value then the goods contained in the transaction will be placed closely. Then transactions that have a high support value will be placed at the end / beginning because the goods in the transaction is the most frequently purchased goods.

REFERENCES

[1] Turban E, Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1, Yogyakarta: Andi, 2005.

[2] Kusrini, Luthfi and Emha Taufiq, Algoritma Data Mining, Yogyakarta: andi, 2009.

[3] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, John Willey & Sons. Inc, 2005.

[4] L. W. Santoso, "Pembuatan Perangkat Lunak Data Mining Untuk Penggalian Kaidah Asosiasi Menggunakan Metode Apriori," *JURNAL INFORMATIKA,* vol. 4, no. 2, pp. 49-56, 2003.

[5] I. Djamaludin and A. Nursikuwagus, "Analisis Pola Pembelian Konsumen Pada Transaksi Penjualan Menggunakan Algoritma Apriori," *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer,* vol. 8, no. 2, p. 671, 2017.

[6] Nurdin and D. Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Pada Supermarket Sejahtera Lhokseumawe," *Techsi,* vol. 6, no. 1, p. 134–155, 2015.

[7] H. N. Wulandari and N. Rahayu, "Pemanfaatan Algoritma Apriori untuk Perancangan Ulang Tata Letak Barang di Toko Busana," in *Seminar Nasional APlikasi Teknologi Informasi (SNATI)*, 2014.

[8] Heroe Santoso dkk, "Data Mining Analisa Pola Pembelian Produk dengan Menggunakan Metode Algoritma Apriori," in *Seminar Nasional Teknologi Informasi dan Multimedia*, Yogyakarta, 2016,.

[9] H. Santoso, I. P. Hariyadi and P. Prayitno, "Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori," *Semnasteknomedia Online,* vol. 4, no. 1, pp. 3-7, 2016.

E. INTERPRETATION / EVALUATION

The results of data mining analysis above can be used for various purposes in sales, one of which is to set the placement of goods or the layout of goods. From the above results can be concluded that in layout arrangements can be done with transactions that have a high confidence value then the goods contained in the transaction will be placed adjacent. Then transactions that have a high support value will be placed at the end / beginning because the goods in the transaction is the most frequently purchased goods.

TABEL V. Final Results For Placement of Goods

No.	Placement of Goods
1	ABC Chili Sauce, ABC Tomato Sauce, Mayonaise, Pepper, Salt, Chili Powder