published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal Homepage: www.idss.iocspublisher.org/index.php/jidss

Implementation of Data Mining to Predict Stocks of Goods Using the Apriori Algorithm at Mom's Kitchen Bakery

Ruminse Situmorang[1], Fricles Ariwisanto Sianturi[2]

Informatics Engineering, STMIK Pelita Nusantara, Jl. Iskandar Muda No. 1 Medan

[1]ruminsesitumorang@gmail.com , [2]sianturifricles@gmail.com
* Corresponding author

ARTICLE INFO

Article history:
Received: 09/08/2020
Revised: 23/08/2020
Accepted: 22/09/2020
Available online 30/09/2020

Key words:
Apriori Algorithm,
Data Mining,
Stock Prediction

ABSTRACT

The application of the Apriori Data Mining Algorithm in predicting stock items can be used to predict what types of goods or brands should be owned or what stocks are right for use in Mom's Kitchen bakery. By utilizing software that is designed or a system that has been made to predict the stock of goods at Mom's Kitchen bakery is one of the right ways to find out what customers are interested in using the Apriori algorithm. The results of this study are, companies can more easily provide products that customers want more based on the stock that has been provided. The system built is web-based using php with MySQL database.

1. Introduction

Data mining is a method used to find data and dig certain hidden data from large data. Data mining can be used to find information on the combination of items in a sale, predict the stock of goods, determine the items to be ordered. In a company that is engaged in sales, it is necessary to increase the company's service sector. An information system is needed that can help companies find out what kind of products should be improved according to customer demand.

Mom's Kitchen is one of the retail stores that is engaged in selling various types of bread, and one of the unique stores in the field of various foods that has reliable employees. CV Mom's Kitchen is active in determining the inventory stock of goods that are seen based on customer demand, it is still experiencing problems, due to the unsupported system. The problem that is often encountered is the difficulty in determining the amount of stock inventory that will be used, this causes stock vacancies and deadcstock for certain items, and often experiences excess and under-stock of goods, experiences errors in inputting stock data, the system used is still in progress. manual form.

Data Mining is a method for finding the best-selling data or the most needed customers. Data Mining has several algorithms, namely: C4.5, K-Means, Support Vector Machines, Apriori, Expectation Maximization Algorithm, PageRank, K-Nearest Neighbors, Naive Bayes, Classification and Regression Trees, Adaboost. The appropriate application to overcome this problem is by developing the concept of data mining. According to Lina Ningsih., Dewi Ayu Nur Wulandari. Data Mining Market Basket Analysis Using Apriori Algorithm to Determine Drug Inventory. National Conference of Social Sciences and Technology, 2017, 3 (3), 227-235. Explaining that inventory has a very important role in increasing sales and service to consumers. If the quantity of goods in stock is more than the quantity demanded,

2. Method

2.1 Research Framework

To assist in the preparation of this research, it is necessary to have a clear framework in stages. This framework represents the steps to be done in solving the problems to be discussed. Following are the steps taken in accordance with the flow of the author's research design as shown in Figure 1

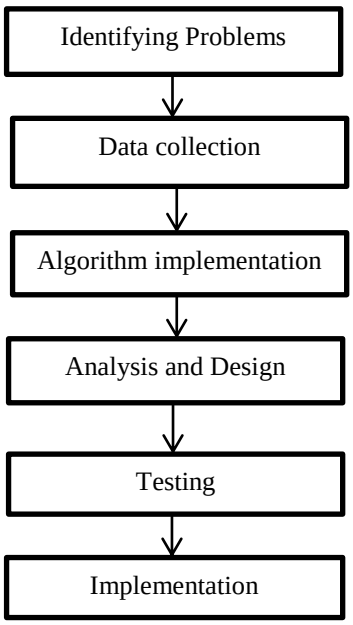

Figure 1. Research Framework

2.2 Data Collection

At this stage the writer will collect data related to the provision of performance allowances with several stages, namely literature study, observation and interviews.

a. Literature review
   The author obtains data by quoting from several reading materials to obtain theoretical data and information that can assist the author in completing this thesis, the author also uses books as the basis for the author's knowledge.
b. Observation
   Is an activity on a process or object by directly observing the provision of performance allowances to obtain clear and correct information.
c. Interview
   Is an oral communication process that aims to gather certain information in a structured way and carried out by two or more people. The interview stage is carried out directly with related parties, in this case with the Head of the General Affairs and Civil Service Subdivision.

2.3 Method of Analysis with the Apriori

Some of the methods that are often mentioned in the literature in the data mining literature include:

a. Classification (classification), classification is a method in data mining that is most often used to solve problems in the real world. Classification is the process of finding a model or function that describes and differentiates data into one of the previously defined classes according to Han and Kamber in Fricles et al (2019: 26)
b. Clustering - Clustering can be used to generate class labels for a group of data. Objects are grouped based on the principle of maximizing intraclass similarities and minimizing similarities between the objects. That is, a group of objects is formed so that the objects in the group have high similarities compared to others, but are somewhat different from objects in other groups.
c. Association (association), the association is also known as Market Basket Analysis. Association analysis is useful for revealing interesting relationships hidden in large datasets.

Apriori Algorithm

The Apriori Algorithm is a basic algorithm that was proposed by agrawal and srikant in 1994 to determine the freuent itemsets for the boolean association rules. A rule that states the

association between several attributes is often called an affinity analysis or market basket analysis. One stage of analysis that has attracted the attention of many researchers to produce an efficient algorithm is the analysis of high frequency patterns (frequent pattern minging). This section aims to analyze and understand the algorithms that will be used in processing data that has been obtained from the research site, especially in the process, in which the a priori algorithm is used.

3. Results and Discussion

3.1 Analysis and application of methods

Analysis is useful for knowing the needs of data mining software needs to be built. In this stage, the search and collection of data and knowledge required to analyze the prediction results of the amount of stock data are carried out. So that in the end the analysis gets the result in the form of a system whose structure can be well and clearly defined. The purpose of system analysis in building data mining is to be able to find out the needs of users and the system, namely regarding the input and output that must be provided by the user and the needs of the system. This process will be an input to the overall system design process. The ability to implement data mining is to solve problems depending on how broad the knowledge base is.The data used is stock data,

Algorithm implementation

The stages of the system a priori algorithm in the data mining process are:

a. Determine the minimum value of support and the value of support
b. Determine the support values for I-itemset, 2-itemset
c. Determine the value of confidence.
d. Establishment of association rules.

Table 1
Goods Stock Data

No.	Date	Name of goods
1	02/10/19	Golden Hana Flour, Eggs, Ceres, Parmesan Cheese, Sugar
2	05/10/19	Butter elle, Coffee, Cocoa powder
3	11/10/19	Whipping cream, Milac, Diamond
4	10/13/19	Maizena, sugar, Hana gold
5	10/15/19	Golden Hana flour, ceres, coffee, eggs, cornstarch
6	18/10/19	Whipping Cream, chakras, Cream Cheess, Cedar cheese
7	22/10/19	Vegetable Oil, Fresh Milk
8	25/10/19	chocoBitz 67%, Cacao Barry 36%, Cacao Barry 50%, Sweetened Condensed Milk, Vanilla essence
9	27/10/19	Cocoa Powder, Whipping Cream, Hana gold flour, Sugar
10	30/10/19	Maizena, Diamond
11	31/10/19	Ceres, Vegetable oil
12	03/11/19	Palmia, Blue Key, Blue Triangle
13	05/11/19	Cream cheess, cornstarch, cranberries, eggs
14	06/11/19	Powdered milk, hana gold flour, millac, diamond, vegetable oil
15	07/11/19	Golden hana flour, granulated sugar, blue keys, blue triangle, eggs
16	08/11/19	Fondant, fondext
17	11/11/19	Butter elle, golden hana flour
18	11/15/19	Maizena
19	17/11/19	Powdered Milk, Chakra
20	20/11/19	Cream cheess, chedar cheese
21	22/11/19	Black sesame, Egg, Yogurt,
22	23/11/19	Golden Hana Flour
23	24/11/19	Mayonnaise
24	28/11/19	Millac, Diamond, Vegetable Oil
25	30/11/19	Sugar, Nestum, Skippy, Wheat, Cranberries, Nutela, Milk Powder, Hanna flour
26	03/12/19	Egg, Fondant, Fondext, Chocolate 58% Tulips, 60% Chocolate Tulip, Colata Chocolate
27	05/12/19	Sugar, Cheese Slice
28	07/12/19	Golden hana flour, Blue Key, Blue triangle
29	10/12/19	Walnut, Blueband, Australian Butter
30	12/15/19	Egg, sticky black sticky rice
31	17/12/19	Butter elle, Maizena, Milk Powder, sugar
32	19/12/19	Maizena, Cream cheess, Golden Hana Flour, Blue key, Matcha

No.	Date	Name of goods
33	21/12/19	White sesame, black sesame, coffee, cocoa powder, pulut flour
34	23/12/19	Tang Mian, Tapioca Flour, Cashew Nuts, Almonds, Whipping Cream, Millac, Diamond, Peanuts, Vegetable Oil, Nestum, Grean Tea Powder, Wheat, Coffee, Cranberries, Nutela, Hana Gold Flour

Given a support value of at least 8% of 34 orders and the support value will be calculated for 1-itemset with the formula:

$$Support\ (A) = \frac{\sum the\ amount\ of\ stock\ data\ contains\ A}{\sum the\ amount\ of\ outgoing\ stock\ data} \text{x}100\%$$

$$\text{Support(golden hana flour)} = \frac{11}{34} \text{x}100\% = 32{,}35\%$$

$$\text{Support(Egg)} = \frac{7}{34} \text{x}100\% = 20{,}59\%$$

$$\text{Support(sugar)} = \frac{7}{34} \text{x}100\% = 20{,}59\%$$

$$\text{Support(Maizena)} = \frac{7}{34} \text{x}100\% = 20{,}59\%$$

Table 2
Category 1 item set

No.	Item1	Qty	Support
1	golden hana flour	11	32.35%
2	Egg	7	20.59%
3	Sugar	7	20.59%
4	Maizena	7	20.59%
5	vegetable oil	5	14.71%
6	Diamond	5	14.71%
7	cream cheess	4	11.76%
8	milk powder	4	11.76%
9	Coffee	4	11.76%
10	blue key	4	11.76%
11	whipping cream	4	11.76%
12	blue triangle	3	8.82%
13	Millac	3	8.82%
14	Ceres	3	8.82%
15	cocoa powder	3	8.82%
16	butter elle	3	8.82%
17	Chakra	2	5.88%
18	Nutela	2	5.88%
19	Cranberry	2	5.88%
20	Wheat	2	5.88%
21	black sesame	2	5.88%
22	Fondext	2	5.88%
23	Fondant	2	5.88%
24	Nestum	2	5.88%
25	chedar cheese	2	5.88%
26	Palmia	1	2.94%
27	brown colata	1	2.94%
28	Peanuts	1	2.94%
29	almond nut	1	2.94%
30	cashew nut	1	2.94%
31	tapioca flour	1	2.94%
32	Tang mian	1	2.94%
33	pulut flour	1	2.94%
34	white sesame	1	2.94%
35	Matcha	1	2.94%
36	black sticky rice	1	2.94%
37	Australian butter	1	2.94%
38	Blueband	1	2.94%
39	Walnut	1	2.94%

From table 2, we will look for the minimum support value≥ 8% of the 110 stock data, data will be obtained that meet the minimum value of support as follows:

Table 3
Calculation Result 1 Item set

No.	Item1	Qty	Support
1	golden hana flour	11	32.35%
2	Sugar	7	20.59%
3	Maizena	7	20.59%
4	Egg	7	20.59%
5	vegetable oil	5	14.71%
6	Diamond	5	14.71%
7	blue key	4	11.76%
8	whipping cream	4	11.76%
9	Coffee	4	11.76%
10	milk powder	4	11.76%
11	cream cheess	4	11.76%
12	blue triangle	3	8.82%
13	Millac	3	8.82%
14	Ceres	3	8.82%
15	cocoa powder	3	8.82%
16	butter elle	3	8.82%

From Table 3, then determine the 2-Itemset candidate from the item stock data by combining the items in pairs.

Previously, we first created a table by combining the monthly stock data, then we could create a 2-itemset candidate. Like the following table

Based on table 3 which contains items with the number they have, then look for the minimum support value≥ 8% as in table 4 then search for the value of the support item with the following formula:

$$Support\ (A,B) = \frac{\sum\ the\ amount\ of\ stock\ data\ contains\ A\ and\ B}{\sum the\ amount\ of\ outgoing\ stock\ data} \text{ x } 100\%$$

$Support$ (golden hana flour $and\ sugar$) $= \frac{4}{34}$ x 100% = 11.76%

$Support$ (golden hana flour $and\ egg$) $= \frac{3}{34}$ x 100% = 8.82%

$Support$ ($blue\ key\ dan\ blue\ triangle$) $= \frac{3}{34}$ x 100% = 8.82%

$Support$ ($Vegetable\ oil\ dan\ Millac$) $= \frac{3}{34}$ x 100% = 8.82%

$Support$ ($Diamond\ dan\ Vegetable\ oil$) $= \frac{3}{34}$ x 100% = 8.82%

$Support$ (golden hana flour and $blue\ key$) $= \frac{3}{34}$ x 100% = 8.82%

$Support$ ($Diamond\ dan\ Millac$) $= \frac{3}{34}$ x 100% = 8.82%

$Support$ ($Egg\ dan\ sugar$) $= \frac{2}{34}$ x 100% = 5.88%

Table 4
2 Set Items

No.	Item 1	Item 2	Qty	Support
1	golden hana flour	Sugar	4	11.76%
2	golden hana flour	Egg	3	8.82%
3	blue key	blue triangle	3	8.82%
4	vegetable oil	Millac	3	8.82%
5	Diamond	vegetable oil	3	8.82%
6	golden hana flour	blue key	3	8.82%
7	Diamond	Millac	3	8.82%
8	Egg	Sugar	2	5.88%
9	Sugar	milk powder	2	5.88%

No.	Item 1	Item 2	Qty	Support
10	golden hana flour	Ceres	2	5.88%
11	Coffee	cocoa powder	2	5.88%
12	whipping cream	Diamond	2	5.88%
13	Sugar	Maizena	2	5.88%
14	cornstarch	cream cheess	2	5.88%
15	Egg	Maizena	2	5.88%
16	Egg	Ceres	2	5.88%
17	golden hana flour	whipping cream	2	5.88%
18	golden hana flour	Millac	2	5.88%
19	golden hana flour	Coffee	2	5.88%
20	golden hana flour	Maizena	2	5.88%
21	golden hana flour	milk powder	2	5.88%
22	golden hana flour	blue triangle	2	5.88%
23	golden hana flour	vegetable oil	2	5.88%
24	golden hana flour	Diamond	2	5.88%
25	Coffee	whipping cream	1	2.94%
26	butter elle	cocoa powder	1	2.94%
27	butter elle	milk powder	1	2.94%
28	butter elle	Maizena	1	2.94%
29	Coffee	Maizena	1	2.94%
30	butter elle	Coffee	1	2.94%
31	Sugar	blue triangle	1	2.94%
32	Coffee	Diamond	1	2.94%
33	cocoa powder	whipping cream	1	2.94%
34	Coffee	vegetable oil	1	2.94%
35	Coffee	Millac	1	2.94%
36	whipping cream	cream cheess	1	2.94%
37	whipping cream	vegetable oil	1	2.94%
38	whipping cream	Millac	1	2.94%
39	Diamond	Maizena	1	2.94%
40	vegetable oil	milk powder	1	2.94%
41	Diamond	milk powder	1	2.94%
42	Maizena	blue key	1	2.94%
43	Maizena	milk powder	1	2.94%
44	cream cheess	blue key	1	2.94%
45	Sugar	blue key	1	2.94%
46	milk powder	Millac	1	2.94%
47	Ceres	Maizena	1	2.94%
48	golden hana flour	cocoa powder	1	2.94%
49	Egg	Coffee	1	2.94%
50	Egg	cream cheess	1	2.94%

From table 4, we will then look for the minimum support value≥ 8% of the 110 stock data, data will be obtained that meet the minimum value of support as follows:

Table 5
Calculation Result 2 Item set

No.	Item1	Item2	Qty	Support
1	golden hana flour	Sugar	4	11.76%
2	blue key	blue triangle	3	8.82%
3	vegetable oil	Millac	3	8.82%
4	Diamond	vegetable oil	3	8.82%
5	golden hana flour	blue key	3	8.82%
6	Diamond	Millac	3	8.82%
7	golden hana flour	Egg	3	8.82%

Furthermore, to find the association rule, steps are taken to calculate the confidence value of each item listed in table 6 based on the following formula:

$$Confidence\ P\ (B/A) = \frac{\sum the\ amount\ of\ stock\ data\ contains\ A\ and\ B}{\sum\sum the\ amount\ of\ outgoing\ stock\ data\ \ A} \text{x} 100\%$$

$$Confidence\ P\ (\text{Diamond dan Millac/Diamond}) = \frac{3}{3} \text{x}\ 100\% = 100\%$$

$Confidence\ P$ (Vegetable oil and Millac/vegetable oil) $= \frac{3}{3}$ x 100% = 100%

$Confidence\ P$ (golden hana flour dan blue key/golden hana flour) $= \frac{3}{4}$ x 100% = 75%

Table 6
Confidence Value Category

No.	Combination of 2 Items		amount	Value of Confidence
1	Diamond	Millac	3	100%
2	Vegetable oil	Millac	3	100%
3	Blue Key	Blue Triangle	3	100%
4	Golden Hana Flour	Blue Key	4	75%
5	Blue Triangle	Blue Key	4	75%
6	Vegetable oil	Diamond	5	60%

Furthermore, if the minimum confidence value is 50%, the association rules that are formed are as follows:≥

Table 7
Confidence Value Category

No.	Combination of 2 Items		amount	Value of Confidence
1	Diamond	Millac	3	100%
2	Vegetable oil	Millac	3	100%
3	Blue Key	Blue Triangle	3	100%
4	Golden Hana Flour	Blue Key	4	75%
5	Blue Triangle	Blue Key	4	75%
6	Vegetable oil	Diamond	5	60%

Association Rule

Based on the steps that have been made, the items that meet the value of support and confidence with a minimum value of confidence ≥ 50%, for that the association rules are formed from table 7 and the following conclusions can be drawn:

a. If you buy Diamond stock data, you will buy Millac stock data with a Support value of 3% and a confidence value of 100%
b. If you buy Vegetable Oil stock data, you will buy Millac stock data with a Support value of% and a confidence value of 100%
c. If you buy Blue Key item stock data, you will buy Blue Triangle stock item data with a Support value of 3% and a confidence value of 100%
d. If you buy gold hana stock data, you will buy Key Biru stock data with a Support value of 4% and a confidence value of 75%
e. If you buy Blue Triangle stock data, you will buy Blue Key item stock data with a Support value of 4% and a confidence value of 75%
f. If you buy Vegetable Oil stock data, you will buy Diamond stock data with a Support value of 5% and a confidence value of 60%

3.2 System Implementation

A. Use Case Diagram

Use case diagram used to describe and analyze the interactions that occur between the user and the system so that they can be understood more easily. The data processing process is only carried out by the admin, while the system will calculate the value.

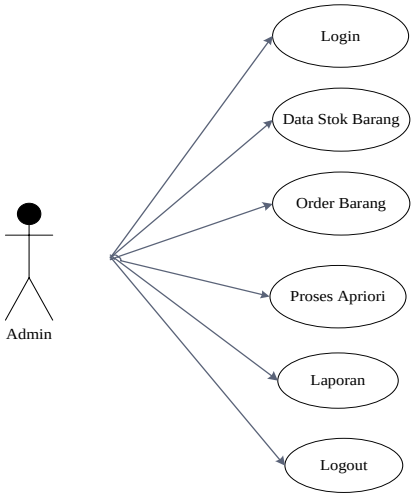

Figure 2. Use case diagram

B. Result Display

This result form is obtained after calculating the 2 item sets and will get the confidence value from highest to lowest with the rule.

C2 (Kandidat 2-itemset)

L2 (Large 2-itemset)

Asosiasi (2-itemset)

No	Rule	Support	Confidence	
1	Jika konsumen membeli **diamond** maka membeli **millac**	8.82%	3/3	100%
2	Jika konsumen membeli **minyak sayur** maka membeli **millac**	8.82%	3/3	100%
3	Jika konsumen membeli **kunci biru** maka membeli **segitiga biru**	8.82%	3/3	100%
4	Jika konsumen membeli **tepung hana emas** maka membeli **kunci biru**	8.82%	3/4	75%
5	Jika konsumen membeli **segitiga biru** maka membeli **kunci biru**	8.82%	3/4	75%
6	Jika konsumen membeli **minyak sayur** maka membeli **diamond**	8.82%	3/5	60%
7	Jika konsumen membeli **diamond** maka membeli **minyak sayur**	8.82%	3/5	60%
8	Jika konsumen membeli **millac** maka membeli **diamond**	8.82%	3/5	60%

Figure 3. Display results

4. Conclusion

Based on the research and implementation carried out on the data mining implementation system in predicting stock items using a priori algorithm, the following conclusions can be drawn:

a. Association analysis with a priori algorithm can find association rules for combinations of stock data, as well as find out what data items are most frequently ordered by employees.
b. The implementation of data mining using a priori algorithm is very efficient and can accelerate the process of forming the trend of itemset combination patterns from the order of stock items at Mom's Kitchen bakery, namely with the highest support and confidence values are Hana Emas and Sugar
c. Testing systems that have been designed to predict stock items on Mom's Kitchen tokoroti using the PHP programming language and MySql database produce a combination of items and percentage values of each transaction data.

5. References

[1] F.A. Hermawati, DATA MINING, I. Yogyakarta: Andi, 2013
[2] F.A.Wowor. F. H & Najoan. X. 2016 Sistem Informasi Perpustakaan Online Berbais Web. E-journal Teknik Elektro dan Komputer,5(2),29-30.

[3] Fricles Ariwisanto Sianturi et al. (2019). Data Mining Teori dan Aplikasi Weka. (Hengki Tamando Sihotang, Ed.).Medan.
[4] Mardiah. (2019). Penerapan Data Mining Apriori Pada Persediaan Obat. Jurnal Ilmu Komputer dan Informatika, 3(2) 121-122.
[5] Pasaribu, S.R. (2014). Implementasi Data Mining Pada Penjualan Penjualan Produk Elektornik Dengan Algoritma Apriori. Pelita Informatika Budi Darma, 119-123
[6] Ramdhany, T. (2019) Implementasi Data Mining Algoritma Apriori Pada Aplikasi Transaksi Penjualan PT. Alfir Untuk Menentukan Posisi Penyimpanan Barang. Jurnal Cloud Information, 4(2) 1.
[7] R. A. S and M. Shalahuddin, Rekayasa Perangkat Lunak. Bandung: Informatika Bandung, 2018.
[8] Dewi, A. & Wulandari, N. 2017,. Data Mining Market Basket Analysis Menggunakan Algoritma Apriori Untuk Menentukan Persediaan Obat. Konferensi Nasional Ilmu Sosial dan Teknologi, 3(3), 227-235.
[9] Santoso, H., Haryadi, I.P., & Prayitno. (2016). Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Algoritma Apriori. Teknik Informatika.(1),19-24.
[10] Sihombing, E.S., Honggowibowo, A.S., & Nugraheny, D. (2012) Implementasi Data Mining Menggunakan Metode Apriori Pada Transaksi Penjualan Barang, 1(1)
[11] Sri, W., Suherman. & Harahap, L.P. (2018), Implementasi Data Mining dalam memprediksi Stok Barang Menggunakan Algoritma Apriori 5(2) 67.
[12] Wijayanti, A. W. (2017). Analisis Hasil Implementasi Data Mining Menggunakan Algoritma Apriori pada Apotek. Jurnal Edukasi Dan Penelitian Informatika (JEPIN), 3(1), 60.