


Information System Data Mining Market Basket Analysis of Purchase Patterns at Mariden Stores Bandar Lampung

Raynaldi Yudhia

Department of Information Systems, Faculty of Computer Science, Institut Informatika Dan Bisnis Darmajaya Bandar Lampung

ARTICLEINFO	ABSTRACT
<p>Article history: Received Dec 30, 2020 Revised Jan 13, 2020 Accepted Jan 22, 2020</p> <hr/> <p>Keywords: Data mining; Association rules; Market basket analysis; a priori.</p>	<p>The tendency of customers to buy goods at the same time is one of the factors in setting the self-service layout to optimally place goods. A strategic and precise product layout in its arrangement will be more easily accessible by consumers and will not waste time. For this reason, it is necessary to create a system model that can determine product layout patterns in supermarkets by seeking the highest confidence. Data mining is one field that is growing rapidly because of the large need for added value from large-scale databases that are increasingly being accumulated. For this reason, the market basket analysis method is used. The market basket is defined as an itemset that is purchased simultaneously by customers in a transaction. This method begins with calculating the Apriori Algorithm to find a number of frequent itemsets and continues with the formation of association rules. The results will produce rules that are very useful for providing information to Mariden stores about the placement of goods according to consumer consumption patterns, providing convenience to consumers in the process of finding goods to be purchased without having to ask the shop owner, being able to see the stock of goods that will run out and can determine the stock of goods because this can actually affect consumer spending tastes and sales of a product.</p> <p><i>This is an open access article under the CC BY-NC license.</i></p> 

Corresponding Author:

Raynaldi Yudhia;
Department of Information Systems Faculty of Computer Science
Darmajaya Institute of Informatics and Business Bandar Lampung
E-mail: raynaldiyudhia@gmail.com

1. INTRODUCTION

Along with the development and business competition in world trade through a free market economy and advances in information technology, companies must be able to implement good business strategies to be able to compete and retain market share. Competition in business cannot be separated from information technology.

In line with the growth of information technology, data mining is one of the rapidly growing fields due to the huge need for added value from large-scale databases that is accumulating more and more. Data mining is a process of exploring the added value that exists in a database by looking at patterns from the data so as to produce useful information that cannot be found manually. Data Mining is also known as Knowledge Discovery in Database (KDD) (Virgiawan and Mukhlash, 2013).

Market basket analysis is generally used as a starting point for seeking knowledge from a data transaction when we do not know what specific pattern we are looking for. The need for market

market basket analysis begins with the accuracy and benefits it produces in the form of association rules. What is meant by association rules are patterns of data linkage in the database.

The tendency of customers to buy goods simultaneously is one of the factors in setting the self-service layout to optimally place goods. A strategic and precise product layout in its arrangement will be more easily accessible by consumers and will not waste time. For this reason, it is necessary to create a system model that can determine product layout patterns in supermarkets by seeking the highest confidence.

2. METHOD

2.1 System Development Methodology

The methodology used in the completion of this research report is the Prototype methodology. The methods in the preparation of this research were carried out with the right technique so that a clear and complete picture of the system to be developed was obtained. Some of the techniques used in this research are as follows:

a. Observation (Observation)

In an effort to collect data and understand the system that is running, the observation technique is the main technique that is commonly and most often used. This technique produces data with a very good level of reliability and accuracy. The observation technique is carried out by observing directly to the object under study so that it can be seen and understood how the system works. In this case the author makes observations at the Mariden Store and consumers who shop there to find out the problems they face so that the author is able to provide some problem solving. This data collection technique was carried out within 2 weeks during the research process.

b. Interview (Interview)

Interview technique is a technique of collecting data by asking questions or asking questions directly. These questions were asked to Mariden Store owners and consumers related to the issues discussed. In this case, the questions and answers were carried out according to the research needs, namely the market basket analysis data mining information system at the Mariden Store.

c. Literature Review (library research)

Literature review is a technique of collecting data by reading, quoting, and collecting theories from books, the internet and studying reference documents and other records that support the research process.

d. Tools and materials

There are several software and hardware used to support the process of implementing the Data Mining Market Basket Analysis Information System, namely Software and Hardware

2.2 Research Framework

The research framework can be seen in the following activity diagrams:

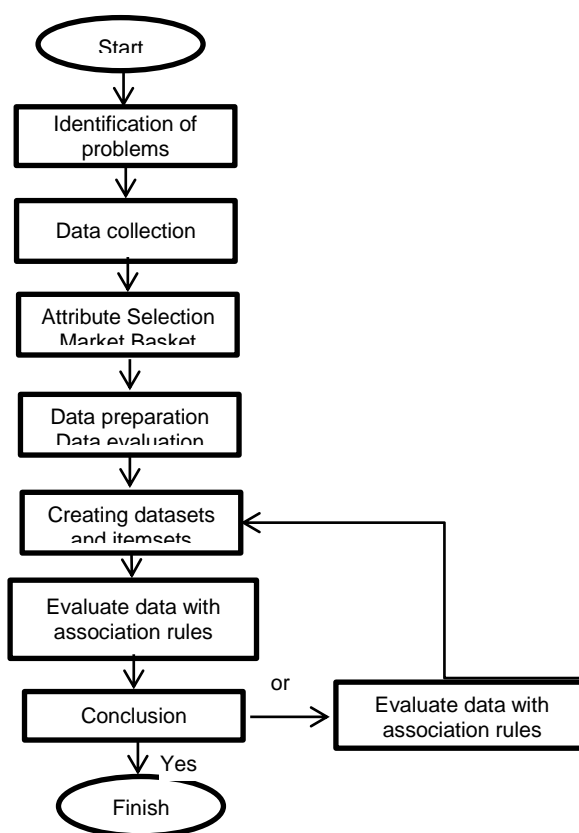


Figure 1. Research Framework

3. RESULTS AND DISCUSSIONS

3.1 Results

The Data Mining process applied in this study uses the Market Basket Analysis (Rule Association) method with the Apriori Algorithm. The steps in the implementation of the Data Mining process are data collection, initialization of data, formation of Association Rules, and drawing conclusions.

1. Apriori Algorithm Search

At this stage, manual calculations are performed using the a priori algorithm by determining the frequent itemset for each transaction. To perform frequent itemset calculations, the minimum support and minimum confidence support values are determined in advance. In this case calculation, the minimum support value is 0.2 (20%) and the minimum confidence support value is 0.1 (10%). The first step that must be done after determining the minimum value of support and minimum confidence is to establish a frequent itemset for 1 itemset.

Table 1.
Formation of frequent itemset for 1 itemset.

Transaction ID Product Name	Bread	Drinking water	Soap	Shampoo	Snack	Cigarette	Instant noodles
1	1	1	0	0	0	0	1
2	1	1	0	1	1	0	0
3	0	0	1	1	0	1	0
4	1	1	1	0	1	0	1
5	0	1	0	0	1	0	0
6	0	0	1	1	0	0	1
7	0	1	0	0	0	1	0
8	0	0	1	1	1	0	3

Transaction ID Product Name	Bread	Drinking water	Soap	Shampoo	Snack	Cigarette	Instant noodles
9	1	1	1	0	1	1	2
10	1	1	0	0	1	0	0
11	1	0	0	0	0	1	0
12	0	0	1	1	0	0	1
13	1	1	0	0	1	0	0
14	0	0	1	1	1	0	0
15	0	0	0	0	0	1	0

After the table above is complete, the next step is to determine the minimum support and minimum confidence support which looks like this:

Table 2.
Determination of minimum support and minimum confidence support

Product name	Number of Transactions Containing IF Products/Total Transactions		
Bread	7	16	0.4375
Drinking water	9	16	0.5625
Soap	7	16	0.4375
Shampoo	6	16	0.375
Snack	8	16	0.5
Cigarette	5	16	0.3125
Instant noodles	6	16	0.375

The table above is a frequent 1-itemset that meets the minimum support value. Where the data above will be used to form a 2-itemset combination pattern starting from the formation of a 2-itemset combination pattern to the formation of a 3-itemset combination pattern. The frequent 2-itemset formation is formed by combining items resulting from frequent 1-itemset calculations that meet the minimum support value that we have determined. The combinations formed from frequent 2-itemsets can be seen in the table below:

Table 3.
Formation of frequent 2-itemset

Product name	Support	
	Number of Transactions Containing IF Products / Total Transactions	
Bread,	0.375	
Water_Drink	0.375	
Water_Drink, Bread	0.125	
Bread, Soap	0.125	
Soap, Bread	0.0625	
Bread, Shampoo	0.0625	
Shampoo, Bread	0.3125	
Bread, Snack	0.3125	
Snack, Bread	0.125	
Bread, Cigarette	0.125	
Cigarettes, Bread	0.1875	
Bread, Instant Noodle	0.1875	
Noodle_Instant, Bread	0.125	
Water_Drink, Soap	0.125	

The frequent 2-itemset calculation table above illustrates the data form of the two-itemset combination. By setting a minimum support value of 0.2(20%), then the items that meet the minimum support value can be seen in the table below:

Table 4.
Formation of frequent 3-itemset

Product name	Support
	Number of Transactions Containing IF Products / Total Transactions
Bread, Water_Drink, Soap	0.125
Water_Drink, Soap, Bread	0.125
Soap, Bread, Water_Drink	0.125
Bread, Water_Drink, Shampoo	0.0625
Water_Drink, Shampoo, Bread	0.0625
Shampoo, Bread, Water_Drink	0.0625
Bread, Water_Drink, Snack	0.3125
Water_Drink, Snack, Bread	0.3125
Snack, Bread, Water_Drink	0.3125
Bread, Water_Drink, Cigarette	0.0625
Water_Drink, Cigarettes, Bread	0.0625
Cigarettes, Bread, Water_Drink	0.0625
Bread, Water_Drink, Noodle_Instant	0.1875
Water_Drink, Instant Noodles, Bread	0.1875
Instant Noodles, Bread, Water_Drink	0.1875
Bread, Soap, Shampoo	0
Soap, Shampoo, Bread	0
Shampoo, Bread, Soap	0
Bread, Soap, Snack	0.125

2. Formation of Rule Association

After the a priori algorithm search is completed, the next step is to find an association rule that meets the minimum confidence requirements. Formation of association rules by calculating using association rules A to B. The strength of association rules is seen by looking at the lift. The table below is a table of candidate Rule Associations consisting of rules generated from frequent itemset searches. From the rules above that meet the minimum confidence value, we can see in the table below:

Table 5.
Rule association candidate

Rule	Support	Confidence	Support x	Fulfil	Fulfil
			Confidence	Tresh Id	Threshold Support x
			Support	Support	Confidence
If Buy Drinking Water Then Buy Drinking Water	0.5625	1	0.5625	Yes	Yes
If Buy Snack Then Buy Snack	0.5	1	0.5	Yes	Yes
If Buy Soap Then Buy Soap	0.4375	1	0.4375	Yes	Yes
If Buy Roti Then Buy Roti	0.4375	1	0.4375	Yes	Yes

This table is a table of candidate Rule Associations consisting of rules generated from frequent itemset searches. From the rules above that meet the minimum confidence value, we can see in the table below:

Table 6.
Rule Association that meets the minimum confidence value

Rule	Support	Confidence	Support x	Meet	Threshold Support x
	Number of Transactions Containing IF Products /Total Transaction	Number of Transactions Containing IF Products / Number of Transactions Containing THEN Products			
If Buy Breadthen BuyDrinking water	0.375	0.66666667	0.25	Yes	Yes
If Buy Air_Minum Then Buy Roti	0.375	0.85714286	0.321428571	Yes	Yes

Rule	Support	Confidence	Support x Confidence	Meet Threshold Support xConfidence	
	Number of Transactions Containing IF Products /Total Transaction	Number of Transactions Containing IF Products / Number of Transactions Containing THEN Products		Yes	Not
If Buy BreadThen Buy Soap	0.125	0.28571429	0.035714286	Yes	Not
If Buy Soap Then Buy Bread	0.125	0.28571429	0.035714286	Yes	Not
If Buy BreadThen Buy Shampoo	0.0625	0.16666667	0.010416667	Not	Not
If Buy Shampoo Then Buy Roti	0.0625	0.14285714	0.008928571	Not	Not
If Buy BreadThen Buy Snack	0.3125	0.625	0.1953125	Yes	Yes
If Buy SnackThen Buy Roti	0.3125	0.71428571	0.223214286	Yes	Yes
If Buy BreadThen Buy Cigarettes	0.125	0.4	0.05	Yes	Yes
If Buy Cigarettes Then Buy Roti	0.125	0.28571429	0.035714286	Yes	Not
If Buy BreadThen Buy Noodle_Instan	0.1875	0.5	0.09375	Yes	Yes
If Buy Noodle_Instan Then Buy Roti	0.1875	0.42857143	0.080357143	Yes	Yes
If Buy Air_Minum Then Buy Soap	0.125	0.28571429	0.035714286	Yes	Not
If Buy Soap Then Buy Air_Minum	0.125	0.22222222	0.027777778	Yes	Not

4. CONCLUSION

The conclusion that can be drawn in this study as described in the previously discussed chapters is that this system makes it easier for shop owners to arrange the layout and find out what items are often purchased at the same time. This system makes it easier for consumers to find goods by knowing the position of the goods without having to ask. The Market Basket Analysis (Rule Association) method with the Apriori Algorithm can be used to help Mariden Stores find out consumer buying patterns and items that are often purchased by consumers.

References

- Andalia, Fanny, Eko Budi Setiawan. 2015. Pengembangan Sistem Informasi Pengolahan Data Pencari Kerja Pada Dinas Sosial Dan Tenaga Kerja Kota Padang. Padang. Vol. 4, No. 2, Oktober 2015
- Asriningtias, Yuli, Rodhyah Mardhiyah. 2014. Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa. Yogyakarta. Jurnal Informatika Vol. 8, No. 1, Januari 2014
- Meilantika, Dian. 2017. Rancang Bangun Sistem Informasi Administrasi Menggunakan Metode Throwaway Prototyping Development Pada Sultan-Sport. Batu Raja. Jutim, Vol 2 No.2, Desember 2017
- Gunadi, Goldie, Dana Indra Sensuse. 2012. Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp- Growth) : Studi Kasus Percetakan Pt. Gramedia. Depok. Jurnal TELEMATIKA MKOM Vol.4 No.1, Maret 2012.
- Masnur, Alkadri. 2015. Analisa Data Mining Menggunakan Market Basket Analysis untuk Mengetahui Pola Beli Konsumen. Riau. SATIN - Sains dan Teknologi Informasi, Vol. 1, No. 2, Desember 2015.
- Mulyanto, Agus. 2009. Sistem Informasi Konsep dan Aplikasi. Yogyakarta: Pustaka Pelajar
- Sugiarto, T. (2012). *Program Pendidikan Anak Usia Dini (PAUD): Suatu Usaha Untuk Menyiapkan Sumber Daya Manusia Berkualitas di Masa Depan.*