

APPLICATION OF DATA MINING ANALYSIS USING APRIORI ALGORITHM FOR BUYING HABITS CUSTOMER FROM INVENTORY DATABASE OF SINAR KARYA FURNITURE

Zein Fikri Muhammad¹

^{1,2}Bachelor of Informatics Engineering, Faculty of Computer Science, Dian Nuswantoro University
Jl. Nakula I No. 5-11 Semarang, Indonesia
Telp. (024) 3517261. Fax: (024) 3520165
E-mail: zeinfikri48@gmail.com¹

Abstract

In this era, furniture company compete with each other tightly. Furniture company usually has several unique and interesting products. A few of products purchased by the customer in cash. The transaction receipt of customers will be saved into kind of database and the total data is so huge. Therefore, needed to analyze those kind of data using appropriate approach. Market basket analysis is one the most data analysis that often use in marketing world that purpose to determine what products are most often purchased at the same time by the customers. This study using apriori algorithm to find frequent itemsets based on sales transaction data. Apriori algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 for the determination of frequent itemset for boolean association rules. It proceeds by identifying the frequent itemsets in the database. One of example is collections of item bought by customers. This study will produce customers buying habits with association rules using apriori algorithm based on the inventory database in Sinar Karya Furniture. The result of confidence value is the customer who buy meja makan, 88.66% also buy kursi makan. And the customer who buy kursi makan, 78.9% also buy meja makan. It is done by using application software development which takes advantages of association rules with apriori algorithm.

Keywords: market basket analysis, association rules, apriori algorithm, buying habits, sinar karya furniture

1. INTRODUCTION

In this global economic era, enterprises compete for their business, especially in world markets. Enterprises should improve their strategy through the use of analysis, technology, skills, and especially the craftsmen to make the greatest enterprises. Called business, because there are customers that interact and make some relationship with the company. In his book [1] argued that companies need to shift their production policy, from the old world of mass product where “standardized products, homogeneous products, and long product life and development cycles were the rule” to the new world where “variety and customization supplant standardized

products”. Pine argued that building one product is simply not adequate anymore. In other words, companies need to be able to develop multiple products that meet multiple customers' needs. While e-commerce hasn't necessarily allowed businesses to produce more products, it has allowed them to provide customer with more choices. By increasing choice, however, it has also increased the amount of information that customers must process before they are able to select which items to meet their need. To address this information overload, e-commerce stores are applying mass customization principles not to the products but to their presentation in the one-line store [1]. The good way to persistent patterns in order to

offer related goods together and track related sales on different levels with market basket analysis.

Market basket analysis is an important component of the analytical system in retail organizations. The widely definition is market basket analysis targets customer baskets in order to monitor buying patterns and improve customer satisfaction.

Apriori is the algorithm that powerful and famous in finding frequent itemsets with using association rule technique. Apriori algorithm was proposed by Agrawal and Srikant 1994 as an algorithm for mining frequent itemsets for Boolean association rules. This algorithm works on the prior knowledge of frequent itemsets properties [2]. Apriori property is used to improve the efficiency of the level wise generation of frequent itemsets. Another algorithm that can used to analyze sales transaction data is Frequent Pattern-Growth (FP-Growth). Like Apriori, the FP-Growth algorithm starts by counting the number of items that appear in dataset.

Problems the availability of transaction data in my enterprises hasn't been utilizing yet for making purpose. On the other hand, several method that can be used to determine frequent itemsets including association rule. In this study, the transaction data will we used to determine frequent itemsets took from "Sinar Karya Furniture" will be used for finding customers' shopping patterns to support making the program.

2. THEORETICAL FOUNDATION

2.1 Market Basket Analysis Definition

Market basket analysis is an important component of the analytical system in retail organizations. There are several definitions of market basket analysis in many researches. As widely meaning, market basket analysis targets customer baskets in order to monitoring buying

patterns and improve customer satisfaction [5]. Another definition, market basket analysis, is one the most data analysis that often used in the marketing world. The purpose of the Market Basket Analysis is to determine what products are most often purchased or used at the same time by the consumer. Market Basket Analysis process is to analyze the buying habits of consumers to find associations between the different products that put consumers in a shopping basket.

2.2 Association Rules

Association rule mining is data mining techniques to find the associative rules between combinations of items. The rule that states the association between some these attributes often called affinity analysis or Market Basket Analysis [5]. Important or not an associative rules can be determined by two parameters, support, and confidence. Support is the percentage of the combination of these items in the database, while confidence is the strong relationships between items in the association rules.

Generally, association rule mining can be described as two phase:

1. Find all of the item combination that meets the minimum requirements of the support value in the database. The value support an item is obtained by the equation below:

$$\text{Support (A)} = \frac{\sum \text{transaction that contains A}}{\sum \text{transaction}}$$

Meanwhile, support value from 2 items are obtained from 2 equations below:

$$\text{Support (A, B)} = P(A \cap B)$$

$$\text{Support (A)} = \frac{\sum \text{transaction that contains A and B}}{\sum \text{transaction}}$$

2. Develop strong association rules from the frequent itemsets, these rules must satisfy minimum support and minimum confidence.

$$\text{Confidence } P(B|A) = \frac{\sum \text{transaction that contains A and B}}{\sum \text{transaction that contains A}}$$

2.3 Apriori Algorithm

Apriori algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 for the determination of frequent itemsets for boolean association rules [2].

Apriori algorithm is used in searching frequent itemset to get the association rules. As the name, this algorithm uses frequent itemset prior knowledge about the properties that have been previously known to process more information. Apriori using an iterative approach, also known as level-wise engines where k-itemset used to find (k + 1)-itemset. First of all searchable set of frequent 1-itemset, this set is denoted as L1. L1 is large itemset used to find L2, then the set of frequent 2-itemset is used to locate L3, and so on until there are no more frequent k-itemset can be found. A Large itemset is frequent or itemset-itemsets that have passed the minimum threshold of support that has been determined [4]. There are two main processes that performed in the apriori algorithm:

1. Join
In this process for each item is combined with another item until can't form a combination anymore. In this step is find the itemsets candidate of 1 item (C1) and count the support
2. Prune
In this process, the result of the itemsets that have been combined was trimmed using a minimum support that has been specified by the user [8].

3. ANALYSIS OF RESULTS AND DISCUSSION

3.1 Application

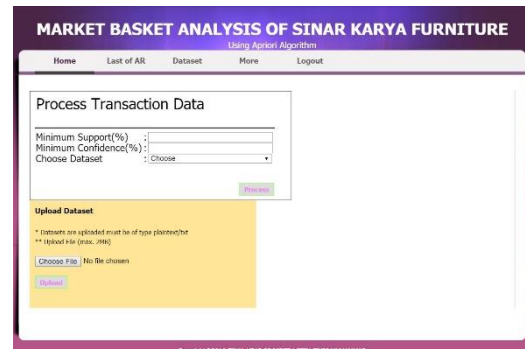


Figure 1. Homepage Apriori

This homepage apriori is the page that used to generates the association rules which relationship each items that purchased simultaneously. The user needs to determine the minimum support and minimum confidence and also choose the dataset that has been uploaded by the admin before in database.

3.2 Analysis and Discussion

From transaction data in Sinar Karya Furniture within January 2012 until July 2012 has a number of the transaction as much as 500. It would be set the minimum support value by 20% and minimum confidence value by 50%.

From all of the transactions, each item will be counted the support value. For calculating the percentage of support value, use the equation below:

$$\text{Support}(A) = \frac{\sum \text{transaction that contains } A}{\sum \text{transaction}}$$

To determine the support value of “Bufet”, the calculation is:

$$\text{Support (Bufet)} = \frac{\sum \text{transaction that contains Bufet}}{\sum \text{transaction}}$$

$$= \frac{122}{500}$$

$$= 24.4 \times 100\% = 24.4\%$$

The items in the last process that meets the minimum support. Next, form the combination item then will be separated by 2 part with each position antecedent and consequent toward all of possibilities. The will be counted again the value support.

Table 1. Large Last-Itemset

Itemset	Support
Kursi makan, meja makan	17.2%

For determining support value is using equation below:

$$\begin{aligned} \text{Support (Kursi makan, meja makan)} &= \frac{\sum \text{transaction(kursi makan, meja makan)}}{\sum \text{transaction}} \\ &= \frac{86}{500} \times 100\% \\ &= 17.2\% \end{aligned}$$

Antecedent → Consequent

Antecedent is the trigger item so that the other item was purchased while consequent is the item is affected by the purchased item antecedent. At the time of generating association rules, the parameters of minimum confidence is needed, because for each association rule that appears will be calculated the percentage value of it confidence according to equation below:

$$\text{Confidence } P(B|A) = \frac{\sum \text{Support A and B}}{\text{Support A}}$$

For example, one member of the set large-itemset is “kursi makan, meja makan” will be established association rules. Then the possibility of rule that will appear is “kursi makan, meja makan” and “meja makan, kursi makan”. Although the support value for {kursi makan, meja makan} and {meja makan, kursi makan} is the same, because the members which arrange is same, this is not applied for association rules. Association rules are implications or unidirectional, therefore “kursi makan → meja makan” and “meja makan → kursi makan” is not same. To calculate the percentage of confidence value from the rule “kursi makan → meja makan”, then the calculation is:

$$\begin{aligned} \text{Confidence (kursi makan} \rightarrow \text{meja makan)} &= \frac{\text{support (kursi makan, meja makan)}}{\text{support (kursi makan)}} \\ &= \frac{17.2}{21.8} = 78.9\% \end{aligned}$$

While for the rule “meja makan → kursi makan”, the percentage of confidence value is:

$$\begin{aligned} \text{Confidence (meja makan} \rightarrow \text{kursi makan)} &= \frac{\text{support (meja makan, kursi makan)}}{\text{support (meja makan)}} \\ &= \frac{17.2}{19.4} = 88.66\% \end{aligned}$$

The rules “kursi makan → meja makan” has confidence value 78.9% and rules “meja makan → kursi makan” has confidence value 88.66%. It indicates that the rule “meja makan → kursi makan” are stronger than “kursi makan → meja makan” because of rules “meja makan → kursi makan” has a bigger confidence value.

The rules “meja makan→kursi makan” will be read “meja makan determining kursi makan”, meja makan is the antecedent which attract the customers to buy kursi makan, while kursi makan is consequent which the item that are affected or purchased when customers decide to buy meja makan. The confidence value is 88.66%, it means that from all of customer (100%) who buy meja makan, 88.66% also buy kursi makan. And the minimum support to get the rules is 10%, so it will appear the association rule.

Table 2. The rules that fulfill the minimum support

Rule(A → B)	Support(A∩ B)	Support(A)	Confidence
kursi makan → meja makan	17.2%	21.8%	78.9%
meja makan → kursi makan	17.2%	19.4%	88.66%

4. CONCLUSION

Apriori algorithm is the one of method that can be applied to sales transaction data which is used to help Sinar Karya Furniture in analyzing costumer purchasing patterns as a decision support the inventory, so Sinar Karya Furniture will be not be in limited supply of consumer demand.

From the analysis and experiment result that have been done, so the researcher summarizes that data mining method is market basket analysis using apriori algorithm that can be applied in the transaction data for determine the promotion in Sinar Karya Furniture with association rules that produced is:

1. meja makan → kursi makan, with the confidence value 88.66% it means that 88.66% from all of the customers that buy meja makan also buy kuris makan.
2. kursi makan → meja makan, with the confidence value 78.9% it means that 78.9% from all of the customers that buy kursi makan also buy meja makan.

With the confidence value above the

author wants to prove the relationship between items. Because the confidence value is the probability of occurrence some products which purchased simultaneously where one of the product is certainly purchased by customer.

REFERENCES

[1] J. B. Schafer, J. A. Konstan dan J. Riedl, “E-Commerce Recommendation Applications,” *Kluwer Academic Publishers. Manufactured in The Netherlands*, no. 5, pp. 116-153, 2001.

[2] P. Tengah, “Penerapan Association Rule Mining Dengan Algoritma A-Priori Untuk Rekomendasi Peminjaman Buku Pada Perpustakaan Daerah Provinsi Jawa Tengah,” eprints.dinus.ac.id, Semarang.

[3] K. Wisaeng, “Association Rule with Frequent Pattern Growth Algorithm for Frequent Item Sets Mining,” *Mathematical Sciences*, vol. VIII, no. 98, pp. 4877-4885, 2014.

[4] H. D. Anggraeni, S. Ragil Saputra dan S. M. Beta Noranita, “Aplikasi Data Mining Analisis Data Transaksi Penjualan Obat Menggunakan Algoritma Apriori,” *Journal of Informatics and Technology*, vol. II, no. 2, pp. 22-28, 2013.

[5] K. Adewole, A. Akintola dan A. Ajiboye, “Frequent Pattern and Association Rule Mining from Inventory Database Using Apriori Algorithm,” *African Journal of Computing & ICT*, vol. 7, no. 3, pp. 35-42, 2014.

[6] J. Han dan M. Kamber, *Data Mining Concepts and Techniques Second Edition*, San Fransisco: Morgan Kaufmann, 2006.

[7] E. N. Sari, "Analisa Algoritma Apriori Untuk Menentukan Merek Pakaian Yang Paling Diminati Pada Mode Fashion Group Medan," *Pelita Informatika Budi Darma*, vol. IV, no. 3, pp. 35-39, 2013.

[8] S. R. Pasaribu, "Sistem Pendukung Keputusan Analisa Pola Penjualan Barang Dengan Algoritma Apriori," *Pelita Informatika Budi Darma*, vol. VI, no. 2, pp. 119-123, 2014.

[9] F. Gorunescu, *Data Mining Concepts, Models and Techniques*, Springer, 2011.

[10] Y. H. Cho dan J. K. Kim, "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, no. 26, pp. 233-246, 2004.

[11] D. T. Larose, *Discovering Knowledge In Data*, Canada: John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[12] S. Maskuroh, "Analisis Keranjang Belanja Dengan Aturan Asosiasi Menggunakan Algoritma Apriori Untuk Stok Barang Pada Listrikmart," *Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro*, Semarang, 2014.