Analysis of Transaction Patterns at Drug Store with Apriori Algorithm

Haryo Kusumo1, Dian Marlina2, Mega Novita3, Muchamad Taufiq Anwar4

1Department of Informatics Management, Stekom University, Jl. Majapahit 605, Kec. Pedurungan, Semarang, Central Java 50192 Indonesia

2Faculty of Pharmacy, Setia Budi University, Jl. Lt. Gen. Sutoyo, Mojosongo, Kec. Jebres, Surakarta City, Central Java 57127, Indonesia

3Faculty of Engineering and Informatics, PGRI University Semarang, Jl. Sidodadi-Timur No.24 Semarang, Central Java 50232, Indonesia

4Faculty of Information Technology, Stikubank University, Jl. Trilomba Juang No. 1 Semarang, Central Java 50241, Indonesia

haryo@stekom.ac.id

**Abstract**. Data mining is a method for finding hidden data from big data which has been continuously applied in various fields such as marketing, education, bioinformatics and so on. Drug store is one of the business sectors that might take the advantage of the data mining. In the drug store, there is a sales transaction data which contains a big number of data. However, there is a limited number of analysis based on this sales transaction data. There are several information that can be obtained from this big data; one of them is the combination of items that consumers often buy. Apriori Algorithm is a data mining method that has been widely used in order to determine the combinations of frequently purchased products. By using the Apriori Algorithm in the sales transaction data of the drug store.

1. Introduction

Much of the available data is often only treated as records without further processing so that it is of no value for future use[1]. Data mining is a method for finding hidden data from big data that is continuously being applied in various fields such as marketing, education, bioinformatics and so on[2]. Drugstores are a business sector that can take advantage of data mining. In pharmacies, there are sales transaction data that contains large amounts of data. However, the analysis based on this sales transaction data is limited[3]. There is some information that can be obtained from this big data; one of which is a combination of goods that consumers often buy[4]. There are many methods in data mining, one of which is association.. The Apriori Algorithm is a data mining method that is widely used to determine product combinations that are frequently purchased . By using the Apriori Algorithm in drugstore sales transaction data, it can help us understand consumer behavior to improve marketing strategies and inventory control.

1. Method

A priori algorithm is an algorithm that is often used to find files that link items with other items and the pattern frequency is high. The a priori algorithm is divided into several stages called iterations[5] [6]:

1. Determine the minimum support
2. The itemset candidate formation, the itemset candidate is formed from a combination of the itemset obtained from the previous iteration. One of the characteristics of the Apriori algorithm is the trimming of candidate itemset subsets of items that contain not included in high frequency patterns.
3. Calculation of support for each itemet candidate.Support from each itemet candidate is obtained by scanning the database to calculate the number of transactions containing all items in the k-itemset candidate. This is also a characteristic of the Apriori algorithm where the computation is required to scan the entire database as the longest itemset.
4. Set high frequency pattern. High frequency patterns include itemset items or sets of candidate itemset
5. Perform the process for the next iteration until no itemset meets minimum support.

The support rule is used to express the proportion of associations in all exchanges made [7] [8];

The supporting formula is as follows[9]:

Meanwhile, when looking for the support value of two items[10]:

To determine the association rules that are formed, at least the itemset must have two candidates A and B. The rules that are formed apply the associative law A -> B does not apply B -> A. To determine the rules A-> B, the formula is used[11]:

1. Result and discussion

In this study, raw data was collected consisting of 12,000 sales transactions of several types of drugs. From the collected data, conversion is carried out in tabular form which can be seen in Table 1 to facilitate data processing

Table 1. Tabular Format Table of Drug Sales Transactions



The application used in this testing process is RapidMiner studio version 9.2. The testing process using the Rapid Miner application consists of several stages, namely (1) add data which is the collection of tabular data on drug sales transactions that have been stored in excel format; (2) selecting data, the process of selecting data is carried out to check the data to be used; (3) the Apriori Algorithm process is a process of combining data with operations related to the Apriori Algorithm; (4) running results, the process of seeing the final results of the application of the Apriori Algorithm in the Rapid Miner Application.

*3.1 Add Data*

Add data which is tabular data retrieval of drug sales transactions that have been saved in excel format. Add data can be done by clicking import data, and looking for the required data in the repository, it can be seen in Figure 1

*3.2 Selecting Data*

The process of selecting data is carried out to cross-check the data to be used in data processing, the selected data is tabular data on drug sales transactions including drug names and transaction frequency, can be seen in Figure 2

*3.3 Apriori Algorithm Process*

After going through the process of adding data and selecting data, the next step is to apply the Apriori Algorithm process to Rapid Miner by dragging and dropping the repository and operator tabs. The process of combining tabular data and the Apriori Algorithm function operator can be seen in Figure 3.

*3.4 Apriori Algorithm Running Results*

At this stage an analysis of the application of the Apriori Algorithm will be generated with drug sales transaction data, the results of the analysis can be seen in Figure 4



Figure 1. Add Data



Figure 2. Selecting Data

Figure 3. The Apriori Algorithm Process



Figure 4. Running Apriori Algorithm Results

1. Conclusion

From the process of calculating drug sales transactions using the Apriori Algorithm, the resulting trends in drug purchases are:

1. If you buy Asamefenamat you will buy amoxillin with 62% confidence
2. If you buy Amixilin, you will buy Asamefenamat with 55% confidence
3. If you buy Cefradoxil, you will buy amoxillin with 61% confidence
4. If you buy Asamefenamat and Sanmol, you will buy Amoxilin with a confidence of 52%
5. If you buy Amoxilin and Anastan, you will buy Asamefenamat with 51% confidence
6. If you buy Asamefenamat and Redoxon, you will buy Amoxilin with 50% confidence

Based on the test results using the Apriori Algorithm with a predetermined minimum value of support and minimum confidence, a drug purchase transaction pattern is produced based on the tendency to purchase drugs purchased by consumers. The results of this test will make it easier for employees to arrange the placement of drug displays as a promotional strategy and also make it easier to monitor drug supplies.

**References**

[1] Han J, Kamber M and Pei J 2012 *Data Mining: Concepts and Techniques*

[2] Sowmya R and Suneetha K R 2017 Data Mining with Big Data *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*

[3] Witten I H, Frank E, Hall M A and Pal C J 2016 *Data Mining: Practical Machine Learning Tools and Techniques*

[4] Wu X, Kumar V, Ross Q J, Ghosh J, Yang Q, Motoda H, McLachlan G J, Ng A, Liu B, Yu P S, Zhou Z H, Steinbach M, Hand D J and Steinberg D 2008 Top 10 algorithms in data mining *Knowl. Inf. Syst.*

[5] Inokuchi A, Washio T and Motoda H 2000 An apriori-based algorithm for mining frequent substructures from graph data *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*

[6] Yu S 2010 A kind of improved algorithm for weighted Apriori and application to data mining *ICCSE 2010 - 5th International Conference on Computer Science and Education, Final Program and Book of Abstracts*

[7] Oliver R L and Swan J E 1989 Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approach *J. Mark.*

[8] Vulcano G, Van Ryzin G and Ratliff R 2012 Estimating primary demand for substitutable products from sales transaction data *Oper. Res.*

[9] Zahay D, Peltier J, Schultz D E and Griffin A 2004 The role of transactional versus relational data in IMC programs: Bringing customer data together *J. Advert. Res.*

[10] Fleischmann M, Kuik R and Dekker R 2002 Controlling inventories with stochastic item returns: A basic model *Eur. J. Oper. Res.*

[11] Musalem E P and Dekker R 2005 Controlling inventories in a supply chain: A case study *International Journal of Production Economics*