Analysis of transaction patterns at drug store with Apriori Algorithm

H Kusumo1*, D Marlina2, M Novita3 and M T Anwar4

1Department of Informatics Management, Universitas Stekom, Jl. Majapahit 605, Kec. Pedurungan, Semarang, Central Java 50192 Indonesia
2Faculty of Pharmacy, Universitas Setia Budi, Jl. Letjen Sutoyo, Mojosongo, Kec. Jebres, Kota Surakarta, Central Java 57127, Indonesia
3Faculty of Engineering and Informatics, Universitas PGRI Semarang, Jl. Sidoarjo-Timur No.24 Semarang, Central Java 50232, Indonesia
4Faculty of Information Technology, Universitas Stikubank, 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 [1] so that it is of no value for future use [2]. Data mining is a method for finding hidden data from big data [3,4] that is continuously being applied in various fields such as marketing, education, bioinformatics and so on [5,6]. Drugstores are a business sector that can take advantage of data mining [7]. there are sales transaction data that contains large amounts of data [8]. However, the analysis based on this sales transaction data is limited [9]. There is some information that can be obtained from this big data; one of which is a combination of goods that consumers often buy [10]. 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 [11], it can help us understand consumer behavior to improve marketing strategies [12,13] and inventory control [14].
2. Methods

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 [15,16]:

- Determine the minimum support.
- The itemset candidate formation, itemset candidate is shaped from a combination of the itemset gotten from the past cycle. 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.
- Calculation of back for each itemset candidate. Support from each itemset candidate is gotten by checking the database to calculate the number of exchanges containing all things within the itemset candidate. Typically moreover a characteristic of the Apriori calculation where the computation is required to filter the whole database as the longest itemset.
- Set high frequency pattern. High frequency patterns include itemset items or sets of candidate itemset.
- 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. The supporting formula is as follows [17]:

\[
\text{Support}(A) = \frac{\text{Number Transaction Contain } A}{\text{Transaction Value}} \tag{1}
\]

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

\[
\text{Support}(A, B) = \text{Support}(A \cap B) = \frac{\text{Number Transaction Contain } A \text{ and } B}{\text{Transaction Value}} \tag{2}
\]

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 \(\rightarrow\) B does not apply B \(\rightarrow\) A. To determine the rules A \(\rightarrow\) B, the formula is used:

\[
\text{confidence} = \text{Support}(A|B) = \frac{\text{Number Transaction Contain } A \text{ and } B}{\text{Number Transaction Contain } A} \tag{3}
\]

3. Results 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 [18] which can be seen in Table 1 to facilitate data processing.

| Amoxilin | Anastam | Asam mefenamat | CDR | Redoxon | Cefradoxil | Kaditic | Sanmol | Sangobion | Metronidazole |
|----------|---------|----------------|-----|---------|------------|---------|--------|-----------|---------------|
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 1       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 1              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 1              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 1       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 1       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 1       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 1       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 1       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 1       | 0          | 0       | 0       | 0         | 0             |
| 1        | 1       | 0              | 1   | 0       | 0          | 0       | 0       | 0         | 0             |
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.

Figure 3. The Apriori Algorithm process.

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 4. Running Apriori Algorithm results.

4. Conclusion

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

- If you buy Asamefenamat you will buy amoxillin with 62% confidence
- If you buy Amixilin, you will buy Asamefenamat with 55% confidence
- If you buy Cefradoxil, you will buy amoxillin with 61% confidence
- If you buy Asamefenamat and Sanmol, you will buy Amoxilin with a confidence of 52%
- If you buy Amoxilin and Anastan, you will buy Asamefenamat with 51% confidence
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 The Morgan Kaufmann Series in Data Management Systems 5(4) 83-124
[2] Sowmya R and Suneeha K R 2017 Data Mining with Big Data Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017 pp 246-250
[3] Witten I H, Frank E, Hall M A and Pal C J 2016 Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations Acm Sigmod Record 31(1) 76-77
[4] Wu X, Kumar V, Ross J Q, 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. 14(1) 1-37
[5] Inokuchi A, Washio T and Motoda H 2000 An apriori-based algorithm for mining frequent substructures from graph data European conference on principles of data mining and knowledge discovery pp 13-23
[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 pp 507-510
[7] Oliver R L and Swan J E 1989 Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approach J. Mark. 53(2) 21-35
[8] Vulcano G, Van Ryzin G and Ratliff R 2012 Estimating primary demand for substitutable products from sales transaction data Oper. Res. 60(2) 313-334
[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. 44(1) 3-18
[10] Fleischmann M, Kuik R and Dekker R 2002 Controlling inventories with stochastic item returns: A basic model Eur. J. Oper. Res. 138(1) 63-75
[11] Chen H, Fuller S S, Friedman C and Hersh W (Eds.) 2006 Medical informatics: knowledge management and data mining in biomedicine (Berlin, Germany: Springer Science & Business Media) vol 8
[12] Peter J P and Olson J C 2009 Consumer Behavior & Marketing Strategy (NY, United States: McGraw-Hill)
[13] Stansfield M 2004 Internet marketing: strategy, implementation and practice Int. J. Inf. Manage. 24(1) 108-110
[14] Musalem E P and Dekker R 2005 Controlling inventories in a supply chain: A case study International Journal of Production Economics 93 179-188
[15] Hegland M 2007 The Apriori Algorithm – A Tutorial Mathematics and computation in imaging science and information processing pp 209-262
[16] Bhandari A, Gupta A and Das D 2015 Improvised apriori algorithm using frequent pattern tree for real time applications in data mining Procedia Computer Science 46 644-651
[17] Kusumo H, Sediyono E and Marwata M 2019 Analisis Algoritma Apriori untuk Mendukung Strategi Promosi Perguruan Tinggi Walisongo J. Inf. Technol. 1(1) 51-62
[18] Kusumo H and Rakasiwi S 2020 Sistem Pendukung Keputusan Rekrutmen Pada Perusahaan Outsourcing Berbasis Web J. Ilm. Ekon. dan Bisnis 13(1) 74-81