Looking for Transaction Data Pattern Using Apriori Algorithm with Association Rule Method

To cite this article: Y Sutisnawati and M Reski 2019 IOP Conf. Ser.: Mater. Sci. Eng. 662 022078

View the article online for updates and enhancements.
Looking for Transaction Data Pattern Using Apriori Algorithm with Association Rule Method

Y Sutisnawati¹, M Reski²
¹Departemen Keuangan dan Perbankan, Universitas Komputer Indonesia, Indonesia
²Departemen Teknik dan Ilmu Komputer, Universitas Komputer Indonesia, Indonesia

Email: reskimuhammad287@mahasiswa.unikom.ac.id

Abstract. The purpose of this research is to find patterns in transaction data so that it can be useful information. The author used data mining in the analysis of transaction data with the association rule method and a priori algorithm. The results of this study state that making the right decision for the marketing strategy. This study will analyze transaction data at XYZ dining places and find relationships between items purchased by consumers.

1. Introduction

In the field of business transactions which occur every time, almost all customers choose more than one type of food. Looking at this pattern, researchers try to do research by processing transaction data that settles and extracts information such as which menu pairs are often ordered. This study uses data mining which is a process that uses mathematical techniques, statistics, machine learning, and artificial intelligence in extracting and identifying useful information from a set of data. [1] The algorithm most often used in data mining in the form of transaction data is the Apriori algorithm. The concept of the algorithm is to collect transaction data by consumers from time to time so that set items that often appear can be generated. The frequency of itemset that has been determined by the user can be found very easily because of the combinatorial explosion. The next step is to make association rules with a confidence that is greater or equal to the specified minimum confidence. [2]

This a priori algorithm is often used to analyze the form of market basketball such as the research of Tang, K., Chen, YL, and Hu, HW (2008) which analyzes purchase patterns at a very detailed concept of time and place, such as a combination of days and stores, in a more general level, such as a combination of quarter and state, and a detailed level combination of one with another general level, such as a combination of days and regions [3]. Raorane A.A.1 Kulkarni R.V.2 and Jitkar B.D.3 also conduct large amounts of data analysis to identify consumer behavior and make effective decisions in achieving competitive advantage over competitors [4]. Srinivas Doddi, Achla Marathe, S. S. Ravi, David C. Torney analyzed medical record data sets using association rules with the aim of obtaining association rules that show the relationship between patient procedures and reports of patient diagnosis [5]. Meiwati, L., and Mustikasari, M. (2010) conducted a study of market basket analysis on drug sales transaction data. [6] LuísCavique also conducted market basket analysis. [7] Maskuroh, S. (2014) also conducted a study on market basket analysis using a priori algorithms on electricity goods stock. [8]

From these studies it can be concluded that the use of a priori algorithms on transaction data at XYZ dining places is very suitable. The purpose of this study is to find the patterns contained in transaction data at XYZ dining places using data mining. The method used is a rule association and a
priori algorithm for transaction data that will look for relationships between items purchased by consumers. It is hoped that this method can help owners determine their marketing strategy decision. The results of this study state that making the right decision for the marketing strategy requires this. This study will analyze transaction data at XYZ dining and find relationships between items purchased by consumers.

2. Method

This study used the data by asking directly to the XYZ Restaurant. Data was obtained in the form of transaction reports along with details of items purchased at XYZ restaurants with a total of 157 transactions. Next is the processing Preprocessing data where attributes that are not used in this study are omitted or deleted according to the needs of this study such as date, price, amount, total and location. The data needed is the item name and the number of items purchased.

Next is to implement a priori algorithm. The first thing to do in the a priori algorithm is to calculate the frequency of items in a transaction, before a minimum has been determined from the high frequency pattern support value in the parable table of 3. Then delete the item that has a frequency below the specified minimum support value. Then the second iteration is done by pairing the remaining items. Then the frequency is searched again and again to prune the item pairs that have a frequency below the minimum support. [9,10]

The following is an explanation table of a priori algorithm that starts from selecting items that have a frequency of at least 3 and merging 2 items then seeing items with frequencies less than 3 removed. Merge the remaining 3 items and get 1 combination that has a frequency of not less than 3 (Figure 1).

| Item | Frekuensi |
|------|-----------|
| I1   | 7         |
| I2   | 8         |
| I3   | 6         |
| I4   | 2         |
| I5   | 3         |
| I6   | 1         |

| Item | Frekuensi |
|------|-----------|
| I1   | 1         |
| I2   | 2         |
| I3   | 1         |
| I5   | 1         |

| Large 1 item | Item | Frekuensi |
|--------------|------|-----------|
| I1           | I2   | 5         |
| I1           | I3   | 4         |
| I1           | I5   | 3         |
| I2           | I3   | 4         |
| I2           | I5   | 3         |
| I3           | I5   | 1         |

| Item | Frekuensi |
|------|-----------|
| I1, I2, I5 | 3         |
| I1, I2, I3 | 2         |

Figure 1. Algorithm Explanation.
3. Results and Discussion

3.1 Data Analysis

Data analysis is done using transaction data. The first is to determine the attributes needed in this analysis. The attribute used is only the item name and the frequency of purchasing each item. From 157 transaction data, there are 32 items with the smallest frequency 2 and the largest is 68 (Table 1).

| No. | Item Name           | Frequency |
|-----|---------------------|-----------|
| 1   | Wedang Susu         | 15        |
| 2   | Teh Manis           | 28        |
| 3   | Milo                | 24        |
| 4   | Telur Puyuh         | 25        |
| 5   | Nasi                | 55        |
| 6   | Bakso Ikan          | 68        |
| 7   | Lemon tea           | 37        |
| 8   | RB Coklat           | 8         |
| 9   | Ayam                | 20        |
| 10  | Bakso Sapi          | 14        |
| 11  | Telur Dadar         | 16        |
| 12  | Tahu                | 30        |
| 13  | RB Milo             | 2         |
| 14  | Cappucino           | 8         |
| 15  | Mie Kuah            | 10        |
| 16  | Mie Goreng KSP      | 5         |
| 17  | Kopi Tubruk         | 12        |
| 18  | Tempe               | 22        |
| 19  | Mie Goreng          | 11        |
| 20  | Air Botol Mineral   | 6         |
| 21  | Kopi Item           | 5         |
| 22  | Sosis               | 10        |
| 23  | RB Telur Daging     | 3         |
| 24  | RB Coklat Keju      | 10        |
| 25  | Scallop Ikan        | 12        |
| 26  | Kerupuk             | 4         |
| 27  | Chocolate           | 11        |
| 28  | RB Stawberry        | 3         |
| 29  | RB Red Velvet       | 6         |
| 30  | RB Green Tea        | 7         |
| 31  | Pisang              | 4         |
| 32  | Bakso Udang         | 6         |
From the a priori algorithm calculation that has been carried out on transaction data at XYZ Restaurant using 157 transaction data, the writer uses a minimum confidence of 65% and produced 2 pairs of items that are often ordered by customers and can be a reference for XYZ Restaurants to make decisions with results relations between items are lost. The following is the following a priori results table (Table 2).

**Table 2. Apriori Results**

| If x Then y          | Total | Support(%) | Confidence(%) |
|----------------------|-------|------------|---------------|
| Teh Manis-Nasi       | 15    | 9,554140127| 53,57         |
| Nasi-Teh Manis       | 15    | 9,554140127| 27,27         |
| Teh Manis-Bakso      | 15    | 9,554140127| 53,57         |
| Bakso Ikan-The Manis | 15    | 9,554140127| 22,06         |
| Teh Manis-Tahu       | 9     | 5,732484076| 32,14         |
| Tahu-The Manis       | 9     | 5,732484076| 30,00         |
| Nasi-Bakso Ikan      | 36    | 22,92993631| 65,45         |
| Bakso Ikan-Nasi      | 36    | 22,92993631| 52,94         |
| Nasi-Tahu            | 19    | 12,10191083| 34,54545455   |
| Tahu-Nasi            | 19    | 12,10191083| 63,33333333   |
| Tahu-Bakso Ikan      | 20    | 12,7388535 | 66,66666667   |
| Bakso Ikan-Tahu      | 20    | 12,7388535 | 29,41176471   |

### 3.2 System Design

After doing the data mining analysis, the next step is to design the system to help data mining process easier to use. The following is the DFD from the system design that will be recommended (Figure 2).

![DFD Level 0](Figure 2)

Level 0 DFD describes the data flow from the recommended data mining process. The user performs minimum input support and minimum configuration and the system will process transaction data in the database so as to produce association rules (Figure 3).
Search $k$-itemset
the process of forming association rules
user
minconf
minsup
Data Transaction
result
Data base
$K$-itemset

**Figure 3. DFD Level 1**

Figure 3 is DFD level 1 where the data flow process is more detailed than the previous level 0 dfd. The user inputs $m_{ins}$ and the system processes the $k$-itemset search on the transaction data contained in the data base. Then the minconf is inputted and the system processes the formation of association rules. After that, the system displays the results. Based on the data flow above. The following is the initial display of data mining applications (Figure 4).

![Data Mining Application Display](image)

**Figure 4. Application Display**

When opening the application for the first time, the application automatically displays transaction data. Users can input minimum support and the specified minimum confidence. Then the user clicks on the process to see the results and if you want to occupy the minimum support and minimum confidance, the user can click reset and click the process again to do the process and the results are displayed.
4. Conclusion
Based on the research that has been done, it can be concluded that there are 2 pairs of items that have more than or equal to 65% confidence which means that these items are more ordered by XYZ restaurant customers. With this analysis, it can help the owner in determining the decision like making a menu package program on these two items or making a package with with 3 items in which 2 items have a confidence value of more or equal to 65% and 1 item that has little frequency or items that are rarely purchased by customers.

Acknowledgements
The researchers intend to express their gratitude to the Chancellor of the Universitas Komputer Indonesia Assoc. Prof. Dr. Ir. Eddy Soeryanto Soegoto, M.T, for helping to provide the material and the information so that this journal can be completed properly.

References
[1] Musen, M. A., Middleton, B., and Greenes, R. A. (2014). Clinical decision-support systems. In Biomedical informatics (pp. 643-674). Springer, London.
[2] 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.
[3] Tang, K., Chen, Y. L., and Hu, H. W. (2008). Context-based market basket analysis in a multiple-store environment. Decision Support Systems, 45(1), 150-163.
[4] Raorane, A. A., Kulkarni, R. V., and Jitkar, B. D. (2012). Association rule–extracting knowledge using market basket analysis. Research Journal of Recent Sciences ISSN, 2277, 2502.
[5] Kamagi, D. H., and Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4. 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. Ultimatics: Jurnal Teknik Informatika, 6(1), 15-20.
[6] Pande, A., and Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Safety science, 47(1), 145-154.
[7] Cavique, L. (2007). A scalable algorithm for the market basket analysis. Journal of Retailing and Consumer Services, 14(6), 400-407.
[8] Fadlil, J., and Mahmudy, W. F. (2007). Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering. Age, 25(26), 25.
[9] Doddi, Achla Marathe, SS Ravi, David C. Torney, S. (2001). Discovery of association rules in medical data. Medical informatics and the Internet in medicine, 26(1), 25-33.
[10] Saifudin, S., Christiyanto, B. F I. (2014). Perancangan Website General Hospital Sebagai Sarana Pemberian Informasi Kesehatan Dan Rumah Sakit. Jurnal Evolusi 2014, Vol 2, No 2.
[11] Soegoto, E. S., and Akbar, R. (2018, August). Effect of the internet in improving business transactions with online market methods. In IOP Conference Series: Materials Science and Engineering, 407(1), p. 012051.