Implementation of Data Mining Sales of Milk Using Apriori Algorithm Method

J Chandra¹, K R Dewi²*

¹Departemen Manajemen, Universitas Komputer Indonesia, Indonesia
²Departemen Sistem Informasi, Universitas Komputer Indonesia, Indonesia

Email: kikirahmawati@mahasiswa.unikom.ac.id

Abstract. The purpose of this research is to find out the items that consumers often buy. The method used in this research is descriptive. The results of this research was to identify the extent of the relationship between one item and another item, and to know the pattern of consumer buying interest in the items purchased. This research was conducted by discussing a combination of items that meet the minimum requirements of the value of support and the formation of association rules that meet the minimum requirements for trust. This research concludes that using the association rule method can help in making decisions about sales strategies.

1. Introduction

Yanbin Ye and Chia-Chu Chiang pointed out that data mining has concerned a lot of helpfulness in the industrial world, especially the trade industry and the people who are exposed to it. One of the main problems in discovering a data is the result of any transaction data from a particular set of items [1]. The use of Data Mining is very cooperative because Data Mining has benefits in business aspects, and to keep away advertising sites for online stores to grow their profits [2] and then with the efficiency and analysis of accurate scheme detection [3]. Data Mining is similarly referred to as (KDD) or knowledge finding in the database [4] and contains integration techniques from various disciplines such as statistics, networks, databases, and information repossess.

Rendering to Dewi Kartika Pane, in the business world especially, the sales there have so many participants competing to get the sales and marketing plans sold by them. Consequently there is a need to increase sales strategies by employing product sales data. The more sales actions are carried out, the more data will be engendered. Advanced data is not only used as company archives, but the data can also be beneficial information to grow sales and product marketing strategies. One field of data mining is the association rule, the task is to treasure subsets of frequently purchased items and the relationships between items by by two main steps: conclusion itemset and generating association rules [5].

According to Christian Borgelt and Rudolf Kruse, the association rule method, or universally known as basket market analysis, aims to find consistency in the shopping behaviour of supermarket customers, online stores, and so on. Through association rules, individual can search for frequently purchased product sets together. The information is articulated in the form of association rules. Association rules can similarly be used to organize products exactly on supermarket shelves or can be placed close to each other to invite more customers to buy them [6]. Apriori [7] is the most
appropriate algorithm for finding item sets that often appear from a transaction. Inappropriately, when the data set size is too large, memory practice and computing budgets are very expensive. Due to the fact that the growth of information continues to produce, the business must deal with the quantity of data that will intensify. Buying is a sequence of actions or doings to obtain properties and services in exchange for their personal use or for resale [8]. Sales are procedures where the needs between shoppers and retailers are rewarded through the discussion of information and their goods [9]. With the existence of strategic planning, the goals and targets of the company become clear so that it will facilitate in formulating other targets and plans and can effectively direct the organization's resources [10]. The aim of this result is to afford outcomes in the form of information that can be beneficial for creating conclusions, preparation of sales strategies, and marketing of dairy products.

2. Method

This research uses the rule method or association rule to find the rules of the relationship amongst arrangement items and use research connected to data mining and apriori algorithms. This method is also used to accumulate datasets and can analyze things that are significant in the application of sales methods and produce approaches. If a customer buys item A and B, he has a requirement of 50% that he will buy item C.

The first thing to do is determine the minimum support. Then the a priori algorithm concept consists of several iteration stages, each iteration will generate a data pattern with the basis of data to obtain support from each item. Items that have support above the minimum support will be selected and become a high frequency pattern called 1 item set. K-item set is a set consisting of k items. In the second iteration process, it will produce 2 items that are determined by each party to have 2 items. The process is continued until there is no minimum support. Support is a large measure that dominates items from the entire transaction. Confidence is a measure that shows the relationship between two items conditionally based on certain conditions [11].

Support value of an item is obtained by formula:

\[
\text{Support (A)} = \frac{\text{The number of transactions containing A}}{\text{Total Transaction}} \tag{1}
\]

While the support value of the 2 items is obtained by the formula:

\[
\text{Support (A} \cap \text{B)} = \frac{\text{The number of transactions containing A and B}}{\text{Total Transaction}} \tag{2}
\]

The Confidence value of rule A → B is obtained by the formula:

\[
\text{Confidence} = P (B \mid A) = \frac{\text{The number of transactions containing A and B}}{\text{Total Transaction Containing A}} \tag{3}
\]
3. Results and Discussion

At this stage it will produce a sales transaction pattern. Transaction data used depends on items purchased by customers. The number of items in the sales transaction data that will be taken is transaction data with the number of items with one or two and only a few items. Furthermore, from the sales transaction, it is formed in the tabular table to make it easier to find out the sales transaction. Samples taken as follows (See Table 1).

| Transaction | M1 | M2 | M3 | M4 | M5 | M6 | M7 |
|-------------|----|----|----|----|----|----|----|
| 1           | 1  | 0  | 1  | 1  | 0  | 0  | 0  |
| 2           | 1  | 1  | 1  | 1  | 1  | 0  | 1  |
| 3           | 0  | 1  | 1  | 0  | 0  | 1  | 1  |
| 4           | 0  | 0  | 1  | 1  | 0  | 0  | 0  |
| 5           | 1  | 0  | 0  | 1  | 1  | 1  | 0  |
| 6           | 0  | 1  | 0  | 0  | 0  | 1  | 0  |
| 7           | 1  | 0  | 0  | 1  | 0  | 0  | 0  |
| 8           | 0  | 0  | 1  | 0  | 0  | 0  | 1  |
| 9           | 1  | 1  | 0  | 1  | 1  | 0  | 0  |
| 10          | 0  | 1  | 1  | 0  | 0  | 0  | 0  |
| 11          | 0  | 0  | 1  | 1  | 1  | 0  | 0  |
| 12          | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| 13          | 0  | 1  | 0  | 1  | 0  | 0  | 0  |
| 14          | 1  | 0  | 1  | 0  | 1  | 0  | 1  |
| 15          | 0  | 1  | 1  | 0  | 0  | 0  | 1  |
| Total       | 7  | 7  | 10 | 7  | 5  | 4  | 5  |

| M1 | M2 | M3 | M4 | M5 | M6 | M7 |
|----|----|----|----|----|----|----|
| 47%| 47%| 67%| 47%| 33%| 27%| 33%|
Product sales transaction patterns from the tabular table results are as follows (see Table 2).

**Table 2. List Transaction**

| Transaction | Itemset         |
|-------------|----------------|
| 1           | M1, M3, M4     |
| 2           | M1, M2, M3, M4, M5, M7 |
| 3           | M2, M3, M6, M7 |
| 4           | M3, M4         |
| 5           | M1, M4, M5, M6 |
| 6           | M2, M6         |
| 7           | M1, M4         |
| 8           | M3, M7         |
| 9           | M1, M2, M3, M5, M6 |
| 10          | M2, M3        |
| 11          | M3, M4, M5    |
| 12          | M1            |
| 13          | M2, M4        |
| 14          | M1, M3, M5, M7 |
| 15          | M2, M3, M7    |

The next footstep is to agree to a minimum limit. Assume the minimum limit is 30%, and then we can regulate the item set frequency. From table 1 it can be opened that items that happened the minimum are \{M1\}, \{M2\}, \{M3\}, \{M4\}, \{M5\}, \{M7\}. Then set item 1 is (see Table 3).

**Table 3. Member Set item 1 for candidate item set 2**

| ITEM SET 1 |
|------------|
|            |
| Item       | Total | Support |
| M1         | 7     | 47%     |
| M2         | 7     | 47%     |
| M3         | 10    | 67%     |
| M4         | 7     | 47%     |
| M5         | 5     | 33%     |
| M7         | 5     | 33%     |

Because from the table upstairs, the support above the minimum support is 30%, so everything goes into the contender of item set 2. Next, estimate the support value in item set 2 (see Table 4).
**Table 4.** Member item set 2 for the specified item is set to 3

| Item | Total | Support |
|------|-------|---------|
| M1   | M1, M2 | 2       | 13%     |
| M2   | M1, M3 | 4       | 27%     |
| M3   | M1, M4 | 4       | 27%     |
| M4   | M1, M5 | 4       | 27%     |
| M5   | M1, M7 | 2       | 13%     |
| M7   | M2, M3 | 5       | 33%     |
|      | M2, M4 | 2       | 13%     |
|      | M2, M5 | 2       | 13%     |
|      | M2, M7 | 3       | 20%     |
|      | M3, M4 | 4       | 27%     |
|      | M3, M5 | 4       | 27%     |
|      | M3, M7 | 5       | 33%     |
|      | M4, M5 | 3       | 20%     |
|      | M4, M7 | 1       | 7%      |
|      | M5, M7 | 2       | 13%     |

From the above table it can be seen that the Support Value that meets the minimum limit is the item (M2, M3) and (M3, M7). Then the item goes to the calculation of item set 3 (see Table 5).

**Table 5.** The item count results set 3

| Item  | Total | Support |
|-------|-------|---------|
| M2    | M3    | M7      | 3       | 20%     |

The table above shows the results where M2, M3, and M7 rose in favour of 20%. Due to the support value not meeting the support of at least 30%, the item set 3 cannot be used to create confidence (see Tables 6 and 7).
Table 6. Results Confidence

| Item      | Total | Support |
|-----------|-------|---------|
| M2, M3    | 5     | 71%     |
| M3, M7    | 5     | 50%     |

Table 7. Results Confidence

| Association Rule | Total | Confidence |
|------------------|-------|------------|
| If you buy M2 it will buy M3 | 5     | 71%        |
| If you buy M3 it will buy M7  | 5     | 50%        |

Accordingly that if you buy M2, it will buy M3, the opportunity is 71% and if you buy M3, it will buy M7 with a 50% accidental. By knowing the results achieved with the most sales, the business can design a sales strategy.

4. Conclusion

The a priori algorithm used in this study aims to provide useful information to determine the pattern of buying milk by buyers with dissimilar types of milk. Association rules are used to find the relationship between item 1 and alternative item based on different types of milk. From the results of the intentions above, the most desirable taste of buyers is M3, M2, and M7. That way for other types can be collective with the type of milk that is frequently purchased. So milk sales are more optimum and sales are consistently distributed.

References

[1] Ye, Y., & Chiang, C. C. (2006, August). A parallel apriori algorithm for frequent itemsets mining. In Software Engineering Research, Management and Applications, 2006. Fourth International Conference on, (pp. 87-94). IEEE.
[2] Chan C C H, Cheng C B, and Hsien W C 2011 Pricing and promotion strategies of an online shop based on customer segmentation and multiple objective decision making. Expert System with Applications, 38(12), pp. 14585-14591.
[3] Oentaryo R J, Lim E P, Finegold M, Lo D, Zhu F, Phua C, and Perera K S 2014 Detecting click fraud in online advertising : a data mining approach. Journal of Machine Learning Research, 15(1), pp. 99-140.
[4] J. Han, M. Kamber,”Data Mining: Concepts and techniques”, Morgan Kaufmann Publishers, Book, 2000.
[5] Borgelt, C., & Kruse, R. (2002). Induction of association rules: Apriori implementation. In Compstat (pp. 395-400). Physica, Heidelberg.

[6] Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems, 83-124.

[7] Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. Advances in knowledge discovery and data mining. Menlo Park, Calif : AAAI Press; 1996.

[8] Berkhin, P. (2006). A survey of clustering data mining techniques. In Grouping multidimensional data (pp. 25-71). Springer, Berlin, Heidelberg.

[9] Susanto, A. (2015). What factors influence the quality of accounting information. International Journal of Applied Business and Economic Research, 13(6), 3995-4014.

[10] Soegoto, E. S., & Rafi, M. S. F. (2018, August). Internet role in improving business transaction. In IOP Conference Series: Materials Science and Engineering, 407(1), p. 012059. IOP Publishing.

[11] Sigit Bahtiar Aji Saputra, Reva Dwiana, Wahyu Dewi Nur Oktaviani, Royana Dwi Isnaeni, Tri Astuti, Nurfaizah. Implementasi Data Mining Algoritme Apriori Pada Penjualan Suku Cadang Motor Delta Motor. Yogyakarta : STMIK AMIKOM Purwokerto.