http://infor.seaninstitute.org/index.php/infokum/index
JURNAL INFOKUM, Volume 8, No.2, JUNI 2020 ISSN 2722-4635

IMPLEMENTATION OF APRIORI ALGORITHM IN DETERMINING THE LEVEL OF PRINTING NEEDS

R. Mahdalena Simanjorang[1], Petti Sijabat[2]
[1,2] Program Studi Teknik Informatika STMIK Pelita Nusantara , Jl Iskandar Muda no 01 sumatera Utara Indonesia
mahdalena95@gmail.com, pettisijabat@gmail.com

Abstract
Competition in the business world, especially in the increasingly difficult printing world, requires developers to find strategies to increase orders for printed products ordered. An increasing number of order data every day can be used to develop marketing strategies if processed correctly. A priori algorithms include the type of association rules in data mining. One of the stages of association analysis that attracts many researchers to produce efficient algorithms is the analysis of high-frequency patterns (frequent pattern mining). The importance of an association can be known by two benchmarks, namely: support and confidence. Support (support value) is the percentage of the combination of these items in the database, while confidence (certainty value) is the strength of the relationship between items in association rules.
Keywords: data mining, ordering, a priori

I. Introduction

With the rapid development of computerized science and technology now it has become an important thing and is needed, especially in the development of organizations and management in every company or business both government and private. The amount of competition in the business world, especially in the printing industry, requires developers to find a strategy that can increase product orders at printing companies. With the printing activity every day, the longer the data will be more and more. Therefore every company must have a good data processing system so that the data generated from these transactions can be useful to be made into a monthly or annual report. The data not only functions as an archive for the company, the data can be used and processed into useful information for increasing product orders.

Printing (printing) is a technology or art that produces copies of an image very quickly, such as words or drawings on paper, cloth, and other surfaces. The development of science and technology is progressing faster, so that at this time the printing industry has become more complete and modern. Printing companies in Indonesia are growing rapidly, both on a large, medium, and small scale. This development also took place in the city of Medan, where there were various printing companies in quite a large number. The development of the number of printing companies has resulted in increasingly high business competition in the printing sector.

2. Literatur Riview

2.1Data Mining

Data mining is a process that employs one or more computer learning techniques (machine learning) to automatically analyze and extract knowledge. Other definitions include induction-based learning (induction-based learning) is the process of forming general concept definitions carried out by observing specific examples of concepts to be learned. Knowledge Discovery in Databases (KDD) is the application of scientific methods to data mining. In this context data mining is one step in the KDD process.

Data mining is defined as a process of finding meaningful relationships, patterns, and new trends by filtering very large data stored in storage, using pattern recognition techniques such as statistical and mathematical techniques.[1]–[4]

2.2. Apriori Algorithm

A priori algorithm is the most famous algorithm for finding high-frequency patterns. High-frequency patterns are patterns of items in a database that have frequencies or support above a certain threshold called the minimum support. A priori algorithm is divided into several stages called iteration or pass, namely:

1. Formation of itemset candidates, k-itemset candidates are formed from a combination (k-1) - itemset obtained from the previous iteration. One way of apriori algorithm is to prune a k-itemset candidate whose subset containing k-1 items is not included in a high-frequency pattern with a length of k-1.
2. Calculation of support for each k-itemset candidate. Support from each k-itemset candidate is obtained by scanning the database to count the number of transactions containing all items in the k-itemset candidate. This is also a feature of the a priori algorithm that requires the calculation by scanning the entire database of the longest k-itemset.
3. Set a high-frequency pattern. High-frequency patterns containing k items or k-item sets are determined from candidate k-itemset whose support is greater than the minimum support.
4. If no new high-frequency pattern is obtained, the whole process is stopped. If not, then k plus one and then go back to section 1. [5]–[7]

2.3 Analysis of High Frequency Patterns with Apriori Algorithms

This stage looks for combinations of items that meet the minimum requirements of the support value in the database. An item's support value is obtained using the following formula:

$$support(A) = \frac{jThe\ amount\ of\ transaction\ containedA}{total\ transactions}$$

Meanwhile, the support value of 2 items is obtained using the formula:

$$suppotr(A,B) = P(A \cap B)$$

$$support(A,B) = \frac{\sum transaction\ contains\ A\ and\ B}{Total\ transactions}$$

Frequent itemset shows items set that have a frequency of occurrence more than the specified minimum value (∅). For example ∅ = 2, then all itemsets that occur more than or twice the same time are called frequent. The set of frequent k-itemset is denoted by Fk. [7]–[9]

2.4 Formation of Association Rules

After all the high frequency patterns have been found, then the association rules are found that meet the minimum requirements for confidence by calculating the confidence of associative rules A → B. The confidence value of the rules A → B is obtained by the following formula:

$$confidence = P(B/A) = \frac{\sum transactioninA\ and\ B}{\sum transaction\ A}$$

To determine the association rules to be chosen, it must be sorted by Support × Confidence. Rules are taken as many as n rules that have the greatest results. [10]-[11]

3. Results and Discussion

3.1 Data requirements

The data used in the implementation of a priori algorithm are outlined in the following pattern:

Table 1. Frequency Patterns

No.	*Itemset*
1	Cake Boxes, Letterhead, Brochures
2	Cake Boxes, Brochures, Letterhead
3	Cake Boxes, Brochures, Letterhead
4	Brochures, Cake Boxes, Name Cards
5	Brochures, Cake Boxes, Name Cards
6	Cake Boxes, Brochures, Letterhead
7	Brochures, Letterhead, Invitations
8	Brochures, Cake Boxes, Envelopes
9	Brochures, Cake Boxes, Letterhead
10	Brochures, Invitations, Envelopes
11	Brochures, Cake Boxes, Letterhead
12	Letterhead, Cake Box, Envelopes

From table 1 above, tabular results are generated in the following table:

Table 2. Tabular Transaction data

No	*A*	*B*	*C*	*D*	*E*	*F*
1	1	0	1	1	0	0
2	1	0	1	1	0	0
3	1	0	1	1	0	0
4	1	0	0	1	1	0
5	1	0	0	1	1	0
6	1	0	1	1	0	0
7	0	1	1	1	0	0
8	1	0	0	1	0	1
9	1	0	1	1	0	0
10	0	1	0	1	0	1
11	1	0	1	1	0	0
12	1	0	1	0	0	1

3.2 Process

1. Formation of the itemset

Process of forming C1 or referred to as 1 itemset with a minimum amount of support = 55%

With the following formula:

$$Support(A) = \frac{\sum transaction\ contains\ A}{\sum transaction} * 100\%$$

The following is the calculation of the formation of 1 itemset:

$$= \frac{\sum transactions\ contain\ Cake\ Boxes}{\sum 12} = \frac{10}{12} * 100\% = 83{,}33\%$$

$$= \frac{\sum transaction\ contains\ Invitation}{\sum 12} = \frac{2}{12} * 100\% = 16{,}67\%$$

$$= \frac{\sum ttransaction\ containing\ Letterhead}{\sum 12} = \frac{8}{12} * 100\% = 66{,}67\%$$

$$= \frac{\sum transaction\ containing\ brochure}{\sum 12} = \frac{11}{12} * 100\% = 91{,}67\%$$

$$= \frac{Transactions\ containing\ business\ cards}{\sum 12} = \frac{2}{12} * 100\% = 16{,}67\%$$

$$= \frac{\sum transaction\ contains\ Envelopes}{\sum 12} = \frac{3}{12} * 100\% = 25\%$$

Based on the description above, it can be made in table 3

Table 3. Support of each item

Itemset	*Support*
Cake Box	83,33%
Invitation	16,67%
Letterhead	66,67%
Brochure	91,67%
Name card	16,67%
Envelope	25%

1. Combination 2 itemset

The process of forming C2 or referred to as 2 itemset with a minimum amount of support = 55%, The following is the calculation of the formation of C2 or 2 itemset:

$$= \frac{\sum transactions\ contain\ Cake\ Boxes\ and\ Invitations}{\sum 12} * 100\%$$

$$= \frac{0}{12} * 100\% = 0$$

$$= \frac{\sum transaksi\ mengandung\ Kotak\ Kue\ dan\ Kop\ Surat}{\sum 12} * 100\%$$

$$= \frac{7}{12} * 100\% = 58{,}33\%$$

$$= \frac{\sum transactions\ contain\ Cake\ Boxes\ and\ Brochures}{\sum 12} * 100\%$$

$$= \frac{9}{12} * 100\% = 75\%$$

$$\frac{transactions\ contain\ Cake\ Boxes\ and\ Business\ Cards}{\sum 12} * 100\%$$

$$= \frac{2}{12} * 100\% = 16{,}67\%$$

$$= \frac{\sum Transactions\ contain\ Cake\ Boxes\ and\ Envelopes}{\sum 12} * 100\%$$

$$= \frac{2}{12} * 100\% = 16{,}67\%$$

$$= \frac{\sum Transaction\ contains\ Invitation\ and\ Letterhead}{\sum 12} * 100\%$$

$$= \frac{1}{12} * 100\% = 8{,}33\%$$

$$= \frac{\sum transactions\ containing\ invitations\ and\ brochures}{\sum 12} * 100\%$$

$$= \frac{2}{12} * 100\% = 16{,}67\%$$

$$= \frac{\sum transactions\ containing\ invitations\ and\ brochures}{\sum 12} * 100\%$$

$$= \frac{0}{12} * 100\% = 0$$

Table 4 2-itemset candidates

Itemset	*value*	*Support*
Cake Box, Invitation	0	0
Cake Box, Letterhead	7	58,33%
Cake Box, Brochure	9	75%
Cake Box, Name Card	2	16,67%
Cake Box, Envelope	2	16,67%
Invitation, Letterhead	1	8,33%
Invitation, Brochure	2	16,67%
Invitation, Business card	0	0
Invitation, Envelope	1	8,33%
Letterhead, Brochure	7	58,33%
Letterhead, Business cards	0	0
Letterhead, Envelopes	1	8,33%
Brochures, Business Cards	2	16,67%
Brochures, Envelopes	2	16,67%
Envelopes, Business Cards	0	0

The minimum support set is 55%, so the combination of 2 itemsset that does not meet the minimum support will be removed, looks like table 5 below:

Table 5 Minimum Support 2 itemset 55%

Itemset	*Support*
Cake Box, Letterhead	58,33%
Cake Box, Brochure	75%
Letterhead, Brochure	58,33%

a. Result

The final association rules are ordered based on the minimum support and minimum confidence that has been determined, can be seen in table 6 below:

Table 6 Final Association Rules

Aturan	Support	Confidence
If ordering a Cake Box, it will order a Brochure	55%	90%

So, based on table 6 above, the most ordered items are Cake Boxes and Brochures. With the knowledge of the most ordered items, the company can find out the most ordered items.

4. Conclusions

The most ordered printed materials in the data provided can be determined by using apriori algorithm, by looking at products that meet the minimum support and minimum confidence, the most ordered items are Cake Boxes and Brochures, but in the calculation of support and configuration, it is difficult if the data which is processed in large quantities.

Reference

[1] S. Džeroski, "Data Mining," in *Encyclopedia of Ecology, Five-Volume Set*, 2008.
[2] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, 2014.
[3] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, 2008.
[4] R. Sowmya and K. R. Suneetha, "Data Mining with Big Data," in *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*, 2017.
[5] C. Borgelt and C. Borgelt, "Efficient Implementations of Apriori and Eclat," *PROC. 1ST IEEE ICDM Work. Freq. ITEM SET Min. IMPLEMENTATIONS (FIMI 2003, MELBOURNE, FL). CEUR Work. Proc. 90*, 2003.
[6] A. Bhandari, A. Gupta, and D. Das, "Improvised apriori algorithm using frequent pattern tree for real time applications in data mining," in *Procedia Computer Science*, 2015.
[7] D. S. Kusumo, M. A. Bijaksana, and D. Darmantoro, "DATA MINING DENGAN ALGORITMA APRIORI PADA RDBMS ORACLE," *TEKTRIKA - J. Penelit. dan Pengemb. Telekomun. Kendali, Komputer, Elektr. dan Elektron.*, 2016.
[8] A. Nursikuwagus and T. Hartono, "IMPLEMENTASI ALGORITMA APRIORI UNTUK ANALISIS PENJUALAN DENGAN BERBASIS WEB," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, 2016.
[9] D. K. Pane, "Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus : Kreditplus)," *Pelita Inform. Budi Darma*, 2013.
[10] H. Toivonen, "Apriori Algorithm," in *Encyclopedia of Machine Learning and Data Mining*, 2017.
[11] J. Nahar, T. Imam, K. S. Tickle, and Y. P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females," *Expert Syst. Appl.*, 2013.