THE OPTIMIZE OF ASSOCIATION RULE METHOD FOR THE BEST BOOK PLACEMENT PATTERNS IN LIBRARY: A MONTHLY TRIAL

T. Husain[1*], **Nuzulul Hidayati[2]**
[1]Department of Information Systems, STMIK Widuri, Jakarta, Indonesia
[2]Faculty of Economics and Business, University of Persada Indonesia Y.A.I, Jakarta, Indonesia

A R T I C L E I N F O	**ABSTRACT**
	Data mining is the process of finding interesting patterns and knowledge from large amounts of data. Sources of information service, especially in the library, include books, reference books, serials, scientific gray literature (newsletters, reports, proceedings, dissertations, theses, and others). The importance of this research being carried out in the library in this study aims to implement data mining with the association rule method to solve problems, especially in the placement of shelves based on the category of the printed version of the book collection. This research method uses a qualitative research approach. Data was collected using documentation techniques and deep analysis of existing weaknesses to identify user needs whose information was obtained through observation and interviews with key informants (admin, user, etc.). For example, the determination of the best book placement patterns can be done by looking at the results of the tendency of visitors to borrow books based on a combination of 2 item sets with 60 percent of confidence value every month or week and must be evaluated or take a calculate again.
Correspondence: T. Husain, Department of Information Systems, STMIK Widuri, Jakarta, Indonesia, Email: thusain050686@gmail.com	

INTRODUCTION

The tendency to use text-based electronically continues to increase from time to time. In addition, the growth of data collected and archived with a computer base contributes to databases that have a large capacity; this requires a technique or method to process the data to become useful information for business decision-making. Therefore, the technology adopted to overcome the above problems is data mining. Data mining is the process of finding interesting patterns and knowledge from large amounts of data [1]. In the new millennium, global libraries are certainly a solution to serve users how to book storage by librarians or users physically in the world of information. However, the uniqueness of the database is that it is a source of information that can be made available to business people to the scientific community in all parts of the world [2]. Furthermore, data plays a central role in helping to understand changing user clearly needs and shaping and restructuring services and procedures suitability in the context of an academic library environment [3]. Sources of information in the library include books, reference books, serials, scientific gray literature (newsletters, reports, proceedings, dissertations, theses, and others). In addition, the library also has digital-based information sources such as e-books, e-journals, audiovisuals, films, photos, slides, transparencies, and others. Most of them demand fast-paced services with complete digital/electronic collections in the context of library users [4]. Therefore, the librarian must have the integrity to provide optimal services following the development of technology and information.

The role of users of this information concerning data mining is inseparable from temporal decision support systems (TDSSs) are interactive decision systems that use temporal data and models to solve problems by considering the dimensions of temporal data, with the domain of knowledge discovery from data (KDD) and human-computer interaction (HCI) [5]. Furthermore, the difference between services, specifically based on traditional libraries and e-libraries, only differs in the media used, namely paper or digital, in accommodating the entire process and service [6].

This research was motivated by previous researchers through the concept of data mining using the algorithm with a particular approach to form a decision tree through the following stages: (i) preparing training data; (ii) determination of the roots of the tree; (iii) determination of gain value; and (iv) repeat the determination of the roots of the tree until all records are perfectly partitioned [7]. Association rule mining with unique specimens such as market-based analysis, which involves 9 customers and 5 item sets categories with letting the minimum confidence threshold is 60 percent, which produces was analyzed to identify potentially interesting patterns to make causal statements about the patterns of purchases according to existing database [8]. The application of data mining to predict wallpaper sales uses the C4.5 Algorithm with the classification method. Utilize 30 assessment data from 6 consumer interest ratings. The findings of this study get the highest factor that affects sales, namely the number of wallpaper motifs. In contrast, other factors such as price, size, material quality, and color do not affect purchasing decisions [9]. Furthermore, using the association rule method for placing books in the library. The above method is obtained based on Apriori algorithm calculations with 10 transactions of borrowing books. The system built can be arranged close to the book's layout to facilitate books that visitors will borrow with minimum confidence, i.e., 83.33 percent [10]. The analysis of the association rule method for product recommendations in online stores. The above method is obtained based on the test results using Apriori knowledge to identify buyers' user preferences and historical data on 20 transactions of 15 items product. The results of the analysis obtained by the association rules can produce the right recommendations with a confidence value of 76.92 percent [11]. Identifying the association rule method aims to the pattern of events' visit of foreign tourists to Bali every month for 37 years transaction. The results obtained by the association rules can produce the right recommendations for a pattern of tourists visiting Bali, with the minimum value of support is 15 percent and the minimum is 60 percent confidence from yield applications [12]. This study aims to merge the success of the Apriori Algorithm from prior research and specifically to determine a trial evaluation for would utilize of monthly period.

One of the libraries belonging to a boarding school in Sukabumi District, West Java, has implemented an online catalog to find information about its collections connected to several computers with a LAN system using one server computer. The online catalog system above is operated using a desktop/laptop computer using software integrated with the SLiMS.web.id official website. However, the implementation of technology so far has not been able to see the results of the tendency of library visitors to the frequency of borrowing books based on the pattern of arrangement of the layout of the book collection. By utilizing the information obtained from library user data, the use of data mining in the case can be used to obtain information about the lending system, the layout of the shelves, and their relationship to obtain optimal book placement results and are easily searchable by library visitors, especially regular visitors within the organization's area. The importance of this research being carried out in the library aims to optimize implement data mining with the Association Rule method to solve best problems, especially in the placement based on the printed version as book collection and the result hope can utilize become the best pattern which evaluated as monthly even weekly by librarians.

LITERATURE REVIEW

Data mining is an activity and not an algorithm or program; ACM Press defines it as the process of discovering new patterns from vast data sets, including methods that are part of artificial intelligence, machine learning, statistics, and database systems [13]. Hereinafter, voted encryption algorithm also involves the asymmetric cryptographic name for a key to doing encryption and decryption [14,15]. Data mining is classified into several tasks/jobs undertaken, i.e., description, estimation, prediction, classification, clustering, and association [16]. In general, the study of data mining includes techniques such as classification, regression, variable selection, clustering, and market basket analysis. Process classification to find a database model can use data mining techniques with the Knowledge Discovery in Databases (KDD) Process Model approach.
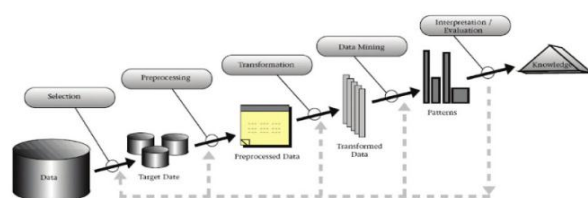

Figure 1. KDD Process Model [5,17]

The KDD process model based on Fayyad in 1995 developed from the following five stages (out of nine activities) as adhere:

1. Data selection consists of two steps. First, develop and understand the application domain. Second, create a target data set from a larger data store. Usually, the data mining maker asks the available data to select relevant data for the detection process. The result of this process is the target data.

2. Data preprocessing, consisting of division with error and missing data. Code values are also ensured to have the same coherent meaning. This stage means that the KDD process model produces the correct results. The result of this process is the data is cleaned or processed.

3. Data transformation, which is the final data processing stage before implementing data analysis techniques. This stage consists of finding useful attributes by applying dimension reduction and transformation methods and finding data representations that do not vary. This stage means that the KDD process model produces the correct results. The result of this process is that the data is transformed.

4. Data mining consists of three steps or sub-tasks. First, select the data mining assignment by matching the objectives defined in the first stage with certain data mining methods such as clustering, regression, classification, etc. Second, choosing a data mining algorithm and selecting methods and parameters to generate data patterns in a certain form of representation. The results of this process are patterns and models.

5. Interpretation/Evaluation consists of two steps. First, interpreting the unearthed pattern, it may be possible to return to the previous stage for further iterations (indicated by the dotted arrow). This stage can also visualize the extracted patterns and models or the extracted data from the extracted models. Second, combining the knowledge extracted by entering it into the performance system or simply documenting and reporting it to the intended party. This process may include checking for and correcting potential errors or discrepancies with previously believed knowledge. This process results from the knowledge that develops from the analysis of existing data with potential knowledge [5].

Apriori algorithm is given by [18] n 1994 for finding frequent itemsets in a dataset for Boolean association rule. Association analysis is one of the data mining techniques used to find interesting relationships between combinations of hidden items in a database. This relationship is represented in the form of association rules [19]. Apriori algorithm is a data retrieval algorithm with associative rules to determine the associative relationship of a combination of items. This algorithm is based on the fact that a priori uses prior knowledge of an item set with frequent occurrences or is called the frequent item set. Apriori uses an iterative approach where the k-item set is used to explore the next (k+1) – item set [1]. The Apriori algorithm scans the database repeatedly to find frequent item sets in forming association rules. Then, the algorithm takes the iteration time is calculated from the time the algorithm ends until it gets the association rule minus the initial time the algorithm is run according to the following formula:

$$t\_old = t\_end - t\_beg$$

Annotation:
t_beg : record the start time of the iteration process
t_end : record the end of time the iteration process

Association rules take the form "If precursor, then the aftermath," throughout with a gauge of the support and confidence associated with the rule. An association rule brings form A, then B (i.e., $A \Rightarrow B$), where the forerunner A and aftermath B are proper subsets of I, and A and B are mutually exclusive. The support 'S' for a particular association rule $A \Rightarrow B$ is the proportion of D transactions containing both A and B calculation like the following formula:

$$Support(I) \quad : \frac{Number\ of\ transactions\ containing\ I}{\sum Number\ of\ transactions} \quad \dots$$

$$(1)$$

The support value of two items is obtained using the following formula:

$$Support(A,B) \quad = P(A \cap B)$$

$$Support(A,B) \quad : \frac{Number\ of\ transactions\ containing\ both\ A\ and\ B}{\sum Number\ of\ transactions}$$

$$\dots (2)$$

The confidence 'c' of the association rule $A \Rightarrow B$ is a gauge of the precision of the rule, as set by the percentage of transactions in D containing A that also contains B as the following formula:

$$confidence = P(A \mid B) \quad = \frac{P(A \cap B)}{P(A)}$$

$$: \frac{Number\ of\ transactions\ containing\ A\ and\ B}{Number\ of\ transactions\ containing\ A} \quad [16]$$

To determine the association rules to be selected, they must be sorted by Support × Confidence. Rules are taken as many as 'n' rules that have the largest result [10], [20].

RESEARCH METHOD

This research method uses a qualitative research approach. Data was collected using documentation techniques and deep analysis of existing weaknesses to identify user needs whose information was obtained through observation and interviews with key informants (admin, user, etc.). However, several procedures still need to be made a decision support model, namely solving problems, especially in placement shelves based on the printed version of the book collection category to make it easier for library visitors to find books within the framework of placing library collections.

The research methodology used and applied in this research includes implementing a concept for library and information science to satisfy the end-users through several phases of data mining as follows [17,21]:

1. The demand of the data by the user's, which requests from visitors to library services in a professional manner in daily activities for users who are involved using a certain approach. This approach can be carried out on different subjects in the form of a top-down or bottom-up approach. In addition, different techniques can obtain data collection to identify problems faced by users.
2. Data mining through disparate warehouses by the library professionals, who library professionals necessaries to recognize the things of the users effectively and systematically.
3. Data filtration and analysis, which data aggregation could be thru the different warehouses.
4. Data export and utilization of data, which the assignment of the library professionals is to analyze and integrate the data for creating beneficial data.
5. Achievement of library goal thru end user's satisfaction, which the commitment of every professional librarian satisfying the request of the users as the destination of assigning the library could be glorious in the evidence scenario nowadays.

Implement Association Rule Mining using Apriori algorithm on borrowing data uses software alternative to perform classification, prediction, cluster, and market basket analysis processes. This software also can normalize the data in an algorithm that may be required via automatic feature selection from wherein each category contribution is proportional to the scale (item set) [21].

RESULTS AND DISCUSSION

Tabulation of data per book item based on 17 days (table 1) as raw data. Hereinafter it aims to find out the yield of book lending activities that book borrowing pattern (table 2) and recapitulate and support each item (table 3).

Table 1. Recapitulation of Borrowing Transaction Data in Class (17-Days)

Recapitulation Date of Transactions	CLASS											
	000 General Books	100 Philosophy | Psychology	200 Religion	300 Social Sciences	400 Language	500 Pure Science	600 Applied Science	700 Art	800 Literature	900 History	813 Fiction Collection	Khusus Reference
13-Feb-2020	1	1	0	1	0	1	1	0	1	0	1	1
14-Feb-2020	1	0	1	0	1	1	0	0	0	1	1	1
15-Feb-2020	0	0	1	0	0	1	1	1	1	1	0	0
16-Feb-2020	0	0	0	0	0	1	1	1	0	1	1	0
17-Feb-2020	0	0	1	1	1	0	1	0	0	1	0	1
18-Feb-2020	0	0	0	1	0	1	0	0	0	0	0	1
19-Feb-2020	0	1	1	0	1	1	1	0	1	1	0	0
20-Feb-2020	1	1	1	0	0	0	1	0	0	1	1	0
21-Feb-2020	0	0	1	0	0	1	0	0	0	1	0	1
22-Feb-2020	0	0	0	0	0	1	1	0	0	0	0	0
23-Feb-2020	0	0	0	1	1	1	0	1	0	1	1	0
24-Feb-2020	0	1	1	0	0	1	0	1	0	1	0	1
25-Feb-2020	1	1	0	1	1	1	1	0	1	1	0	0
26-Feb-2020	0	0	1	0	1	1	1	1	1	0	0	1
27-Feb-2020	1	1	1	0	0	1	1	0	0	0	1	1
28-Feb-2020	0	0	1	0	1	0	1	0	1	1	0	1
29-Feb-2020	0	1	0	0	0	1	0	1	0	1	0	0

Table 1 describes the specification of library visitor needs must be stated in clear technical terms to fulfill the analysis conducted by a librarian. The recapitulation of transaction data groups the types of categories of books that are borrowed the most to determine the pattern of borrowing certain books that are reprocessed in table 2 below:

Table 2. Book Borrowing Pattern

Day 1	General Books	Philosophy | Psychology	Social Sciences	Pure Science	Applied Science	Literature	Fiction Collection	Reference
Day 2	General Books	Religion	Language	Pure Science	Literature	History	Fiction Collection	Reference
Day 3	Religion	Pure Science	Applied Science	Art	Literature	History	-	-
Day 4	Pure Science	Applied Science	Art	History	Fiction Collection	-	-	-
Day 5	Religion	Social Sciences	Language	Pure Science	Applied Science	Reference	-	-
Day 6	Social Sciences	Pure Science	Reference	-	-	-	-	-
Day 7	Philosophy | Psychology	Religion	Language	Pure Science	Applied Science	Literature	Reference	-
Day 8	General Books	Philosophy | Psychology	Religion	Applied Science	History	Fiction Collection	-	-
Day 9	Religion	Pure Science	History	Reference	-	-	-	-
Day 10	Pure Science	Applied Science	Literature	-	-	-	-	-
Day 11	Social Sciences	Language	Pure Science	Art	History	Fiction Collection	-	-
Day 12	Philosophy | Psychology	Religion	Pure Science	Art	History	Reference	-	-
Day 13	General Books	Philosophy | Psychology	Social Sciences	Language	Pure Science	Applied Science	Literature	History
Day 14	Religion	Social Sciences	Pure Science	Applied Science	Fiction Collection	Reference	-	-
Day 15	General Books	Philosophy | Psychology	Religion	Pure Science	Applied Science	Fiction Collection	Reference	-
Day 16	Religion	Language	Applied Science	Literature	History	Reference	-	-
Day 17	Philosophy | Psychology	Pure Science	Art	History	-	-	-	-

Table 2 describes the number of book borrowings made by library visitors, which will be used as data for the needs of the calculation process using the Apriori algorithm. This next stage is looking for a combination of items that meet the minimum requirements of the value of the support in the data. The settling based on the loan data category in Table 3 below is 1 item set. Here's one example of the calculation C1 with item set:

$$Support(Pu\_Science) \quad : \frac{15}{17} \to 88{,}24\%$$

And so until all categories support 1 item set.

Table 3. Support of each 1 item set

Book Category	Total	Support
General Books	5	29,41%
Philosophy | Psychology	7	41,18%
Religion	10	58,82%
Social Sciences	6	35,29%
Language	6	35,29%
Pure Science	15	88,24%
Applied Science	11	64,71%
Art	6	35,29%
Literature	7	41,18%
History	12	70,59%
Fiction Collection	6	35,29%
Reference	9	52,94%

The process of formation of C2 with 2 item sets. Here's one example of calculation 2 item set (from table 2 pattern):

$$Support(PuScie\_AppScie) \quad : \frac{9}{17} \to 52{,}94\%$$

Table 4. Support of 2 item set

Book Category	Total	Support
Pure Science, Applied Science	9	52.94%
Pure Science, Religion	5	29.41%
Pure Science, History	8	47.06%
Pure Science, Reference	8	47.06%
Applied Science, Religion	7	41.18%
Applied Science, History	5	29.41%
Applied Science, Reference	6	35.29%
Religion, History	6	35.29%
Religion, Reference	9	52.94%
History, Reference	4	23.53%

Based on the combination of 2 item sets with minimum support, the combination of 2 item sets that meet the minimum support standard is the book category Pure | Applied Science and Religion | Reference with each support, i.e., 52,94 percent. After one of the patterns of borrowing

frequency can be identified and found, then look for associative rules that meet the minimum requirements for confidence by calculating the confidence of the associative rule A → B. One example of the confidence value of the A → B rule is obtained as follows:

$$Condifence(PuScie\_AppScie) \quad : \frac{9}{15} \to 60\%$$

Table 5. Confidence 2 items

Book Category	Total	Confidence
Pure, Applied Science	9	60%
Pure Science, Religion	5	33,33%
Pure Science, History	8	53,33%
Pure Science, Reference	8	53,33%
Applied Science, Religion	7	46,67%
Applied Science, History	5	33,33%
Applied Science, Reference	6	40%
Religion, History	6	40%
Religion, Reference	9	60%
History, Reference	4	26,67%

Based on manual compute with association rule method from Table 5, the combination of 2 item sets with minimum support of 60 percent yields that the combination of 2 item sets meets the minimum standard of support [8], namely Pure | Applied Science and Religion | Reference classify with support each of 60 percent.

The evaluation stage from a calculation with the Association Rule method produces the best confidence value determined is 60 percent. Therefore, the belief that a value of 1 means the rule must have occurred for 17 days can be enforced, and each month must calculate a book placement pattern with the same technique. If library visitors during that period borrow books in the Pure Science, Applied Science category, the books collection of category arranged in the closest position or front of the library service gates in the one of next month.

CONCLUSION

Determining the pattern of placement books in the library can be done by applying Apriori algorithm with the Association Rule method. The determination of the pattern of book placement can be done by looking at the results of the tendency of visitors to borrow books based on a combination of 2 item sets. This technique is obtained based on the results of Apriori algorithm calculations. The system built can be arranged close to the book's placement to facilitate books that visitors will borrow.

Categories take in this study are the presence or absence of an amount that is only at least one transaction in the borrowing of the diary. Therefore, data sets in evaluating monthly trials should use one full month concern nowadays of daily transactions to produce evaluation results that can be used more accurately next month. The evaluation method can use software testing to make it easier for the user (librarian) to calculate this Apriori Algorithm method.

Hereinafter, proposed calculating can be done every beginning month of time even weekly transactions by future research. This evaluation to obtain the best book placement pattern so that it is easier for always visitors to find.

REFERENCES

[1] Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining Concepts and Techniques*, 3rd Ed. USA: Morgan Kaufmann Publishers, 2015.

[2] Martin Alan Kesselman and Irwin Weintraub, *Global Librarianship*, Allen Kent, Ed. New York • Basel: Marcel Dekker, Inc., 2004.

[3] Faten Hamad, Razan Al-Aamr, Sinaria Abdel Jabbar, and Hussam Fakhuri, "Business Intelligence in Academic Libraries in Jordan: Opportunities and Challenges," *International Federation of Library Associations and Institutions*, vol. 47, no. 1, pp. 37–50, 2021.

[4] Shiksha Kendra, *Library and Information Science: Class XI*, 1st ed., Shri. Vineet Joshi, Ed. Delhi, India: Central Board of Secondary Education, 2014.

[5] Hela Ltifi, Christophe Kolski, Adel M. Alimi, and Mounir Ben Ayed, "A Human-Centred Design Approach for Developing Dynamic Decision Support System Based on Knowledge Discovery in Databases," *Journal of Decision Systems*, vol. 22, no. 2, pp. 69-96, March 2013.

[6] Roopesh K. Dwivedi and R. P. Bajpai, "Use of Data Mining in the field of Library and Information Science: An Overview," in *2nd International CALIBER on 'Road Map to New Generation Libraries Using Emerging Technologies', 11-13 February*, New Delhi, 2004, pp. 512-519.

[7] Prabowo Pudjo Widodo and Rahmadya Trias Handayanto, *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains, 2013.

[8] Irina Tudor, "Association Rule Mining as a Data Mining Technique," *BULETINUL: Universitǎţii Petrol – Gaze din Ploieşti*, vol. LX, no. 1, pp. 49 - 56, 2008.

[9] Juna Eska, "Penerapan Data Mining untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5," *JURTEKSI*, vol. 2, no. 2, pp. 9-13, Maret 2016.

[10] Robi Yanto and Hendra Di Kesuma, "Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule," *JATISI*, vol. 4, no. 1, pp. 1-10, September 2017.

[11] Chavid Syukri Fatoni, E. Utami, and F. W. Wibowo, "Online Store Product Recommendation System Uses Apriori Method," in *International Conference on Electrical, Electronic, Informatics and Vocational Education (ICE-ELINVO 2018)*, vol. 1140, Yogyakarta, 2018, p. 012034.

[12] Dewa Agung Krishna Arimbawa P., I Nyoman Angga Prabawa, and Putu Arya Mertasana, "Implementation of Apriori Algorithm in Determining Tourism Visit Patterns to Bali," *International Journal of Engineering and Emerging Technology*, vol. 4, no. 1, pp. 10-14, 2019.

[13] Sunyoto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika, 2017.

[14] Muneer Bani Yassein, Shadi Aljawarneh, Ethar Qawasmeh, Wail Mardini, and Yaser Khamayseh, "Comprehensive Study of Symmetric Key and Asymmetric Key Encryption Algorithms," in *International Conference on Engineering and Technology (ICET)*, vol. 17615788, Antalya, Turkey, 2017, pp. 1-7.

[15] Taufik Hidayat and Rahutomo Mahardiko, "A Systematic Literature Review Method On AES Algorithm for Data Sharing Encryption On Cloud Computing," *International Journal Of Artificial Intelegence Research*, vol. 4, no. 1, pp. 49-57, 2020.

[16] Daniel T. Larose, *Discovering Knowlege in Data: An Introduction to DATA Mining*. New Jersey: John Wiley & Sons, Inc., 2005.

[17] Bedadyuti Sahoo and B.S.P. Mishra, "Data Mining Is A Perpetual Concept For Library And Information Science: An Estimated Overview," *International Journal of Digital Library Services*, vol. 5, no. 3, pp. 14-21, July-September 2015.

[18] Rakesh Agrawal and Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in *VLDB '94:*

Proceedings of the 20th International Conference on Very Large Data Bases, Santiago de Chile, 1994, pp. 487-499.

[19] Pang-Ning Tan, Michael Steinbach, and Anuj Karpatne, *Introduction to Data Mining*, 1st Ed. USA: Pearson Addison-Wesley, 2006.

[20] Chonny | Microsoft SWE Intern. (2020, October) MEDIUM. [Online]. https://towardsdatascience.com/apriori-association-rule-mining-explanation-and-python-implementation-290b42afdfc6

[21] Galit Shmueli, Nitin R. Patel, and Peter C. Bruce, *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, 2nd Ed. Canada: John Wiley & Sons, Inc, 2010.