

Implementation of a priori algorithm for book lending at state high school library I Silima Pungaga-Punga Parongil

Andini Yulistia Ningsih^{1)*}, Volvo Sihombing²⁾, Sahat Parulian Sitorus³⁾

¹⁾²⁾³⁾Labuhanbatu University, North Sumatra, Indonesia

¹⁾andini150701@gmail.com, ²⁾volvolumbantoran@gmail.com, ³⁾sahatulb@gmail.com

Submitted: Jan 10, 2022 | **Accepted :** Jan 17, 2022 | **Published :** Jan 21, 2022

Abstract:The library plays a role in helping students to enjoy reading books. The availability of books in various fields motivates students to come to visit the library, students / i can read or borrow library books. For this reason, the purpose of this research was carried out including helping library officials apply the rules of how to visit the libraries. In carrying out research methods, including conducting direct observations, conducting interviews to collect the necessary data. The pattern that will be analyzed is the pattern of borrowing what books are often borrowed so that library officials know the information of books that are often borrowed. The result obtained is with the application of a priori algorithms, book data is processed to produce a pattern of borrowing books. After all high frequency patterns are found, then the association rules are sought that meet the minimum requirement for associative confidence $A \rightarrow B$ minimum confidence = 25%. The final association rules are ordered based on minimum support and minimum confidence, if borrowing IPA, then borrowing MTK Support = 15%, Confidence = 42.8%.

Keywords: Data Mining, Lending Patterns, Books, Libraries, Algorithms A priori

INTRODUCTION

Every school has a library. State High School Library 1 Silima Pungga Pungga is one of the libraries that provides a variety of books for students who will read and borrow to get knowledge information. Of the many collections of books in the student library it is difficult to find books to read and borrow or can only be read in place. To facilitate the process of borrowing library books, especially in terms of finding books to be borrowed or students who often borrow books must search for themselves on bookshelves, then there needs to be a proper system in recommending what books are available in the library and that are often borrowed simultaneously by library visitors.

Data mining is the process of finding patterns or interesting information in selected data using certain techniques or methods. One method that is often used in data mining technology is the method of association or association rule mining. Association rule mining is a mining technique for finding association rules between combinations, with regard to the study of "what's with what" (Mardi, 2017).

In this study, the association rule method with a priori algorithm to book lending data in the State High School Library 1 Silima Pungga-Pungga to help implement the book placement system as a loan of books to library members. With this the system of placing books in the library is more effective and efficient. A priori algorithms include a type of association rule in data mining (Sulastris & Gufroni, 2017). In addition to a priori, those included in this group are generalized rule induction methods and hash based algorithms. The rules that state associations between multiple attributes are often called affinity analysis or market basket analysis. Association rule analysis is a data mining technique for finding associative rules between a combination of items. An a priori algorithm that aims to find frequent items is run on a set of data. A priori analysis defines a process for finding all a priori rules that meet the minimum requirements for support and the minimum requirements for confidence (With et al., 2018).

LITERATURE REVIEW

Data mining is a technology that combines traditional analytical methods with sophisticated algorithms to process large volumes of data. Data mining is a term used to find hidden knowledge in a database. Data mining is a semi-automated process that uses statistical, mathematical, intelligence, and machine learning techniques to extract and identify useful and useful information stored in large databases. (Witten et al., 2016)

* Andini Yulistia Ningsih



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Some of the earliest definitions of data mining focused on data processing in the automation process. The theory of data mining is widely used in the field of marketing, sales, and customer service with the same goal of defining data mining as the process of exploration and analysis of large amounts of data automatically and semi-automatically so that meaningful patterns or rules are found. (Wu et al., 2014)

The automated analysis performed by data mining exceeds that performed by traditional decision support systems that are already widely used. Data mining can answer business questions that traditionally require a lot of time and high costs. Data mining explores databases to find hidden patterns, looking for information to predict what businesses and education practitioners might forget because they are beyond their expectations (Fricles Ariwisanto Sianturi, Petti Indrayati Sijabat, Amran Sitohang, 2020).

Rapid developments in data processing technology have made it easier for humans to collect large amounts of data, resulting in a lot of data. Data mining is the process of automatically searching for information that is useful in large data storage. Other terms often used include knowledge discovery (mining) in databases. (KDD) (Sowmya & Suneetha, 2017)

The terms data mining and Knowledge Discovery in Database (KDD) are often used interchangeably to describe the process of extracting hidden information in large databases. Actually the two terms have different concepts, but are related to each other. And one of the stages in the entire KDD process is data mining (Sianturi et al., 2018).

The a priori algorithm is the basic algorithm proposed by Agrawal and Srikant in 1994 for the determination of frequent itemsets for boolean association rules. ((Suyanto, 2017) This algorithm controls the development of itemset candidates from frequent itemsets results with support-based pruning to eliminate unattractive items by assigning minsups. (Esling & Agon, 2012)

Here are the steps of the association rules calculation process (Meilina, 2015)

1. The system scans the database to get a 1-itemset candidate (a set of items consisting of 1 item) and calculates its support value. Then the support value is compared to the minimum support that has been determined, if the value is greater or equal to the minimum support then the itemset are included in the large itemset.
2. Items that are not included in large itemset are not included in the next iteration in prune (in prune).
3. In the second iteration the system will use the results of large itemset in the first iteration (L1) to form the second itemset candidate (L2). In the next iteration the system will use the results of large itemset in the next iteration will use the results of large itemset in the previous iteration (Lk-1) to form the following itemset candidate (Lk). The system will combine Lk-1 with Lk-1 to get Lk, as in previous iterations the system will remove /prune itemset combinations that are not included in large itemset.
4. After the join operation, then the pair of new itemset results of the join process are calculated support.
5. The process of forming candidates consisting of the join and prune process will continue until the candidate set is null, or no more candidates will be formed.
6. After that, from the results of frequent itemset is formed association rules that meet the values of support and confidence that have been determined.
7. In the formation of an association rule, the same value is considered as one value.
8. Association rule formed must meet the minimum value that has been determined.

According (Fricles A Sianturi, Hasugian, Paska Marto, Simangunsong Agustina, 2019) Association analysis is defined as a process for finding all association rules that meet the minimum requirements for support (minimum support) and the minimum requirements for confidence (minimum confidence).

The basic methodology of association analysis is divided into two stages:

1. High Frequency Pattern Analysis

This stage looks for a combination of items that meet the minimum of support values in the database. The support value of an item is obtained by the following formula:

a. *Support*

Support of an association rule is the presentation of a combination of such items in a database, where if you have item A and item B then support is the proportion of transactions in the database that contain A and B (Badrul, 2016). The formula for calculating the support values of the two items is as follows (Febriyani & Winanjaya, 2021)).

The support value of an item is obtained by the following formula

$$\text{Support}(A) = \frac{\text{Number of transactions contained A}}{\text{total transaction}}$$

While the support value of 2 items is obtained from the following formula:

$$\text{Support}(A, B) = \frac{\text{Support}(A, B) = P(A B)}{\sum \text{Transactions Containing A and B}} \div \sum \text{Transaction}$$

* Andini Yulistia Ningsih



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

To find the support value of the 3 items obtained by using the formula:
 $Support(A,B dan C) = P(A \cap B \cap C)$

$$Support(A, B, C) = \frac{\sum Transactions\ Containing\ A, B\ and\ C}{\sum Transaction}$$

b. Confidence

Confidence from the association rule is a measure of the speed of a rule, namely the presentation of transactions in a database containing A and containing B. With confidence, we can measure the strength of the relationship between items in the association rule. The formula for calculating the confidence value of the two items is as follows.

$$Confidence = P(B|A)$$

$$Confidence = \frac{\sum Transactions\ Containing\ A\ and\ B}{\sum Transactions\ Pregnant\ A}$$

To determine the association rules to be selected, they must be sorted by Support x Confidence. Rules are taken as many as n rules that have the largest result.

METHOD

The method used in solving the problem is a field study by conducting a direct field study to collect data. The methods in this research are as follows:

a. Observation

Made a visit to the Silima Pungga Pungga Parongil Public High School Library, namely on Jl. Lamp No. 7 Parongil

b. Interview

Collecting data by conducting direct questions and answers with the librarian of the object under study to obtain data to be studied and data collected based on facts or according to the needs of the Silima Pungga Pungga Parongil Public High School Library. The problems found during the observations were that the library staff did not check the books every day so that the staff did not know the condition of the books in the library, the book lending system still uses manuals, so that the librarian still has difficulty serving students, every student who visits Libraries take a long time to borrow books due to the placement of books that are not in accordance with each subject.

c. Observation

This data collection method is used to obtain data related to the a priori algorithm method to determine the rules for borrowing library books.

d. Literature review

This method is used to obtain additional information that is used as a reference in the development of a priori algorithms.

e. Knowledge Acquisition

The process of collecting knowledge data is done by collecting references from books, articles, papers, journals on data mining and also a priori algorithms.

RESULT

The data taken is loan transaction data in June 2018. The data is data that represents the entire transaction data of 20 transactions. And it can be seen in Table 1.

Table 1
Book Lending Transactions

Visitors	Borrowed book
1	Mathematics, Science, Indonesian
2	Civics, Mathematics
3	IPA, Physical Education
4	Religion, Civics, ICT
5	Science, ICT, Religion
6	Economics, ICT
7	English, Science, Mathematics, ICT
8	ICT, Religion, Physical Education
9	Science, ICT, Religion
10	ICT, Religion

* Andini Yulistia Ningsih



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Visitors	Borrowed book
11	physical education
12	Indonesian, Mathematics, Science
13	IPS
14	IPS, Physical Education
15	Mathematics, English, Indonesian
16	Physical Education, Economics, Social Sciences
17	ICT, English
18	IPA, IPS
19	English, Indonesian
20	ICT

The data in the transactional database above is usually represented in a form as shown in Table 2.

Table 2

Book Lending Transaction Data Representation

Visitors	Book Code	Borrowed book
1	MTK	Mathematics
1	IPA	IPA
1	B. Indo	Indonesian
2	civics	civics
2	MTK	Mathematics
3	IPA	IPA
3	PEN	physical education
4	AGM	Religion
4	civics	civics
4	ICT	ICT
5	IPA	IPA
5	ICT	ICT
5	AGM	Religion
6	EKO	Economy
6	ICT	ICT
7	B. Indo	Indonesian
7	IPA	IPA
7	MTK	Mathematics
7	ICT	ICT
8	ICT	ICT
8	AGM	Religion
8	PEN	physical education
9	IPA	IPA
9	ICT	ICT
9	AGM	Religion
10	ICT	ICT
10	AGM	Religion
11	PEN	physical education
12	B. Ing	English
12	MTK	Mathematics
12	IPA	IPA
13	IPS	IPS
14	IPS	IPS
14	PEN	physical education
15	MTK	Mathematics
15	B. Ing	English
15	B. Indo	Indonesian
16	PEN	physical education
16	EKO	Economy
16	IPS	IPS
17	ICT	ICT

* Andini Yulistia Ningsih



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Visitors	Book Code	Borrowed book
17	B. Ing	English
18	IPA	IPA
18	IPS	IPS
19	B. Ing	English
19	B. Indo	Indonesian
20	ICT	ICT

$$Support(A) = \frac{\text{Number of Transactions Contained A}}{\text{total transaction}}$$

While the support value of the 2 items is obtained from the following 2 formula:

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

$$Support(A, B) = \frac{\sum \text{Transactions Contains A and B}}{\sum \text{Transactions}}$$

Step 1:

Looking for C1 (1-itemset candidate) as follows:

$$Support(Religion) = \frac{\text{Number of transactions contained Religion}}{\text{total transactions}} = \frac{5}{20} = 25\%$$

$$Support(B. Indo) = \frac{\text{Number of transactions contained B. Indo}}{\text{total transactions}} = \frac{4}{20} = 20\%$$

Step 2:

Based on the table containing the items with their support, then look for L1 = { Large 1-itemset } by selecting items that meet the minimum support value of 10% as follows.

$$Support(TIK) = \frac{\text{Number of transactions contained TIK}}{\text{total transactions}} = \frac{9}{20} = 45\%$$

$$Support(IPA) = \frac{\text{Number of transactions contained IPA}}{\text{total transactions}} = \frac{7}{20} = 35\%$$

$$Support(Religion) = \frac{\text{Number of transactions contained Religion}}{\text{total transactions}} = \frac{5}{20} = 25\%$$

Step 3:

The next process is to produce a 2-itemset or C2 candidate by going through the join stage, namely by combining books or items, a candidate item set will be produced. Support count is the number of item combinations in a book borrowing transaction, which means how often the item combination appears or is borrowed.

$$Support(Religion, B. Indo) = \frac{\text{Number of transactions contained Religion, B. Indo}}{\text{total transactions}} = \frac{0}{20} = 0\%$$

$$Support(Religion, B. Ing) = \frac{\text{Number of transactions contained Religion, B. Ing}}{\text{total transactions}} = \frac{0}{20} = 0\%$$

$$Support(Religion, EKO) = \frac{\text{Number of transactions contained Religion, EKO}}{\text{total transactions}} = \frac{0}{20}$$

Step 4:

After calculating and finding support for each 2-itemset candidate, the pruning stage, namely eliminating items that do not meet the minimum support value of 5%, can be generated as follows:

$$Support(Religion, IPA) = \frac{\text{Number of transactions contained Religion, IPA}}{\text{total transactions}} = \frac{1}{20} = 5\%$$

$$Support(Religion, PEN) = \frac{\text{Number of transactions contained Religion, PEN}}{\text{total transactions}} = \frac{1}{20} = 5\%$$

$$Support(Religion, PKn) = \frac{\text{Number of transactions contained Religion, PKn}}{\text{total transactions}} = \frac{1}{20} = 5\%$$

$$Support(Religion, TIK) = \frac{\text{Number of transactions contained Religion, TIK}}{\text{total transactions}} = \frac{4}{20} = 20\%$$

a. Establishment of association rules

After all high-frequency patterns are found, then we look for association rules that meet the minimum requirements for associative confidence $A \rightarrow B$ with minimum confidence = 25%. The confidence value of the rule $A \rightarrow B$ is obtained by the formula

$$Confidence = P(B|A) = \frac{\sum \text{Transaction pregnat A and B}}{\sum \text{Transaction pregnat A}}$$

* Andini Yulistia Ningsih



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The results of the formation of the Association Rule (Association Rule):

$$\begin{aligned}
 \text{Confidence} &= P(\text{IPA}|\text{Religion}) = \frac{\sum \text{Transaction pregnant Religion and IPA}}{\sum \text{Transaction pregnant Religion}} = \frac{1}{5} \\
 &= 20\% \\
 \text{Confidence} &= P(\text{PEN}|\text{Religion}) = \frac{\sum \text{Transaction pregnant Religion and PEN}}{\sum \text{Transaction pregnant Religion}} = \frac{1}{5} \\
 &= 20\% \\
 \text{Confidence} &= P(\text{PKN}|\text{Religion}) = \frac{\sum \text{Transaction pregnant Religion and PKN}}{\sum \text{Transaction pregnant Religion}} = \frac{1}{5} \\
 &= 20\% \\
 \text{Confidence} &= P(\text{TIK}|\text{Religion}) = \frac{\sum \text{Transaction pregnant Religion and TIK}}{\sum \text{Transaction pregnant Religion}} = \frac{4}{5} \\
 &= 80\% \\
 \text{Confidence} &= P(\text{B.Ing}|\text{B.Indo}) = \frac{\sum \text{Transaction pregnant B.Indo and B.Ing}}{\sum \text{Transaction pregnant B.Indo}} = \frac{2}{4} \\
 &= 50\% \\
 \text{Confidence} &= P(\text{IPA}|\text{B.Indo}) = \frac{\sum \text{Transaction pregnant B.Indo and IPA}}{\sum \text{Transaction pregnant B.Indo}} = \frac{2}{4} \\
 &= 50\%
 \end{aligned}$$

Meanwhile, the final association rules based on minimum support and minimum confidence can be seen in Table 3.

Table 3
Final Association Rules

Rule	Support	Confidence
if you borrow religion, then borrow ICT	20%	80%
if you borrow B.Indo, then borrow B.Ing	10%	50%
if borrowing IPA, then borrowing MTK	15%	42.8%

DISCUSSIONS

From the test results with the application used, the calculation results show that:

- If you borrow a religious book, you will borrow a science book
- If you borrow a religious book, you will borrow a physical education book
- In the printing process the association is based on the min support and confidence values entered by the user so that the final result will appear according to the user's wishes. The following can be shown to print the final result of borrowing a book.

The results of testing with the application can be seen in the following image:

* Andini Yulistia Ningsih





Fig.1 Form Book Lending Transactions

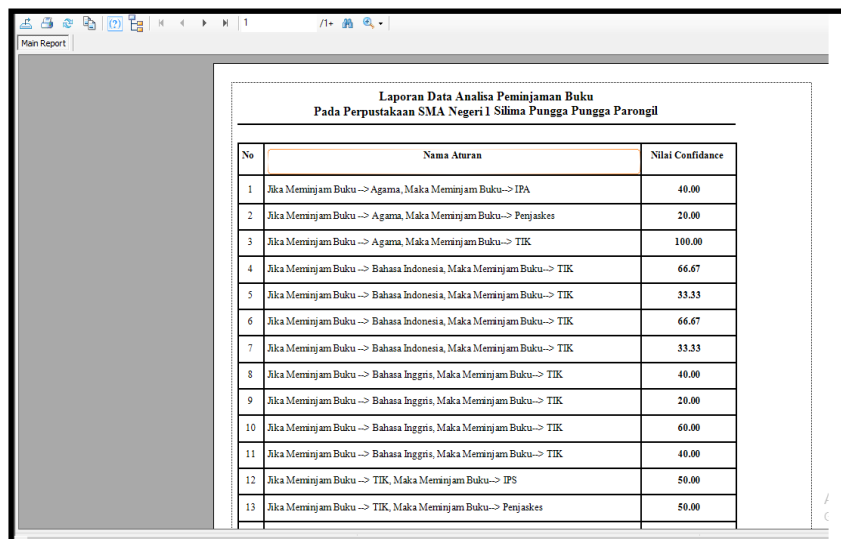


Fig.2 Book Loan Analysis Data Report

CONCLUSION

Based on the analysis and system testing that has been done, it can be concluded as follows:

1. In the system to be built, this research starts from the data collection stage to the final stage which produces book lending analysis data.
2. The association rules using the algorithm are found. From the results of the final confidence value with the association rules, the data pattern meets the minimum requirements.
3. The data used in processing should be grouped by type and publisher of books, because of the rapid development of book publishing.
4. The Data Mining Application Using the Apriori Algorithm Method to Determine the Rules of Book Lending Transactions at the Library of SMA Negeri 1 Silima Pungga Pungga Parongil is expected to continue to be developed according to existing needs.
5. It is hoped that with the Data Mining Application Using the Apriori Algorithm Method to Determine the Rules of Book Lending Transactions at the Library of SMA Negeri 1 Silima Pungga Pungga Parongil this can be utilized by librarian as well as possible.

1. REFERENCES

- Badrul, M. (2016). Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan. *None*.
Dengan, S., Algoritma, C., & Hidayati, W. (2018). *Data Mining Penentuan Tenaga Perawat Di RSUD Sultan*. 1(2), 1-7.

* Andini Yulistia Ningsih



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Esling, P., & Agon, C. (2012). Time-series data mining. *ACM Computing Surveys*. <https://doi.org/10.1145/2379776.2379788>
- Febrivani, E., & Winanjaya, R. (2021). Penerapan Data Mining Asosiasi Pada Persediaan Obat. 3(3), 354–365.
- Fricles A Sianturi, Hasugian, Paska Marto, Simangunsong Agustina, N. B. (2019). Data Mining [Teori dan Aplikasi Weka. In -: Vol. (Issue).
- Fricles Ariwisanto Sianturi, Petti Indrayati Sijabat, Amran Sitohang, R. M. S. (2020). APRIORI ALGORITHM FOR THE DETERMINATION OF THE. *Infokum, Jurnal*, 9(1), 11–20.
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- Meilina, P. (2015). Penerapan Data Mining dengan Metode Klasifikasi. *Jurnal Teknologi Universitas Muhammadiyah Jakarta*.
- Sianturi, F. A., Informatika, T., & Utara, S. (2018). Penerapan Algoritma Apriori Untuk Penentuan Tingkat. *Mantik Penusa*, 2(1), 50–57. <http://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/330>
- Sowmya, R., & Suneetha, K. R. (2017). Data Mining with Big Data. *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*. <https://doi.org/10.1109/ISCO.2017.7855990>
- Sulastri, H., & Gufroni, A. I. (2017). Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Teknologi Dan Sistem Informasi*, 03(02), 299–305. <https://doi.org/10.25077/TEKNOSI.v3i2.2017.299-305>
- Suyanto. (2017). Data mining Untuk Klasifikasi dan Klasterisasi Data. *SpringerReference*.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. In *Data Mining: Practical Machine Learning Tools and Techniques*. <https://doi.org/10.1016/c2009-0-19715-5>
- Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2013.109>