Utilization of Data Mining for Placement of Books in the Library Using the Apriori Method

Rahmat Robi Waliyansyah \*, Mega Novita, Nugroho Dwi Saputro

Faculty of Engineering and Informatics, PGRI University Semarang

rahmat.robi.waliyansyah@upgris.ac.id \*

**Abstract**. The Apriori Algorithm is an algorithm that is well known for performing frequent itemset searches using the association rule technique. The Apriori algorithm uses knowledge about the previously known frequent itemset to process data related to borrowing books in the library to be processed into useful information. This study uses a priori algorithm method for grouping books at the PGRI University Library in Semarang based on trends that appear together in a library activity visit. By gaining knowledge from this algorithm, it can be used as a reference for the PGRI University Library in Semarang in placing book collections.

1. Introduction

College libraries are libraries that are in the environment of colleges or colleges, academies and other higher education, which are essentially an integral part of a college[1]. The library is one of the facilities for providing knowledge and information, the library is also a place for the teaching and learning process for users to get the desired information[2]. With advances in information technology today, the need for accurate information is needed in everyday life, so that information will become an important element in the development of society today and in the future[3]. Book lending transaction data is very important data for a library so that every borrowing transaction in the library, the transaction data will be stored in the data warehouse[4]. Therefore, we need a new information system that is able to analyze book lending transaction data in the library database. The data will be reprocessed into useful information[5]. The development of information technology has contributed to the rapid growth in the amount of data collected and stored in large databases (data stacks)[6]. Libraries that do not have an integrated data warehouse usually manually or use application software that is used to connect the separate modules[7]. By using data mining techniques using a priori algorithms to find data sets that appear most frequently in a data set[8]. By using the a priori algorithm, recommendations will be given to libraries for book placement based on borrowed data[9], thus helping improve services to the library[10].

1. Research materials and method

The object of this research is the library of the PGRI University of Semarang (UPGRIS). The data used by researchers for this research is book lending transaction data from the UPGRIS library. The data used is the transaction data for borrowing books in the library for one full month (September). Analysis of book lending transactions is only carried out with a sample that is already available. The method used for analysis only uses the Apriori method with association rules. Tests were carried out using the Orange Biolab tools.

Apriori is an algorithm that is well known for searching frequent itemset using the association rule technique[11]. The a priori algorithm uses knowledge about the previously known frequent itemset to process further information. This algorithm helps to explore patterns of association in books based on transactions that occur, so that the system can provide recommendations for books to be borrowed by users[12]. The a priori algorithm is used to determine possible candidates by paying attention to minimum support. The support value of an item is obtained using the following formula:

$$Support \left(A\right)=\frac{number of transactions containing A}{total transactions}$$

The support value of 2 items is obtained using the formula:

$$Support \left(A,B\right)=P(A∩B)$$

$$Support \left(A,B\right)= \frac{\sum\_{}^{}transactions containing A and B}{\sum\_{}^{}transactions}$$

The frequent item set includes a set item that has a frequency of occurrence more than the specified minimum value.

The principle of the a priori algorithm is[13] :

1. Collect single item quantity, get big item.
2. Get candidate pairs, count -> large pairs of items.
3. Get candidate triplets, count -> large triplets of items and so on.
4. As a hint: every subset of a frequent itemset must be a frequent.

The next stage is the creation of association rules. This rule creation is done after all high frequency patterns have been found. Search for association rules that meet the minimum confidence requirements by calculating confidence A → B. The confidence value from the rule A → B obtained the following formula:

$$Confidence P\left(A\right)= \frac{\sum\_{}^{}transactions containing A and B}{\sum\_{}^{}transactions containing A} $$

Association rules are determined and selected should be sorted based on support x confidence. N rules are taken that have the largest payoff. A rule that states the association between attributes is often called an affinity analysis or market basket analysis[14].

The process of calculating the association rule consists of several stages, which are as follows:

1. The system scans the database to get a 1-itemset candidate (a set of items consisting of 1 item) and calculates the support value. Then the support value is compared to the minimum support that has been determined, if the value is greater than or equal to the minimum support, the itemset is included in the large itemset.
2. Itemset that are not included in the large itemset are not included in the next iteration (prune).
3. In the second iteration the system will use the results of the large itemset in the first iteration (L1) to form the second candidate itemset (L2). In the next iteration the system will use the results of the large itemset in the next iteration will use the results of the large itemset in the previous iteration (Lk-1) to form the following itemet candidate (Lk). The system will join (join) Lk-1 with Lk-1 to get Lk, as in the previous iteration the system will prune the combination of itemset that is not included in the large itemset.
4. After the join operation is carried out, the new itemset pair resulting from the join process is calculated for support.
5. The candidate formation process, which consists of a join and prune process, will continue until the itemet candidate set is null, or no more candidates will be formed.
6. After that, from the results of the frequent itemset an association rule is formed that meets the predetermined values ​​of support and confidence.
7. In the formation of the association rule, the same value is considered as one value.
8. The association rule that is formed must meet the predetermined minimum value.
9. For each large itemset L, we look for a non-empty subset L for each subset, resulting in a rule of the form aB (La) if its support (L) and support (a) are greater than minimum support.

Before the data is processed by data mining, the data needs to go through the preprocessing stage. This stage relates to the selection and transfer of useless data (data cleaning), merging data sources (data integration), transforming data in a form that can simplify the process (data transformation), displaying data in a more readable amount (data reduction) . Everything comes from raw data (transaction data) and the results will be data that will be ready to be processed by data mining[15].

1. Results and discussion

In this study, a priori algorithm was used to find the connection between books in each transaction which was used as an alternative recommendation for other book choices if the library members wanted to borrow. Before the data is mined, the data first goes through the preparation process so that the mining process will produce recommendations according to the transaction data with the appropriate number of recommendations. The author's transaction data gets from the results of joining several existing tables, then combines the transaction data of each member into weekly transactions.

From the results of the author, it shows that the data on book lending at the PGRI University of Semarang so far has not been well optimized, so that the data on borrowing is increasing day by day. The data is only stored as campus archives or bookkeeping and there is no known benefit from the existing data.

**Table 1.** Example of data for borrowing books for September 2019

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Student NPM** | **Book title** | **Borrow Date** | **Return Date** |
| 1 | 18120255 | Three-Colored Realm | 02-09-2019 | 09-09-2019 |
| 2 | 17320052 | Compilation: Drama at the Edge of the Scalpel (A Note on Imperfect Science) | 02-09-2019 | 09-09-2019 |
| 3 | 18150023 | Statistics for Research | 02-09-2019 | 09-09-2019 |
| 4 | 18150023 | Descriptive Statistics for Education | 02-09-2019 | 09-09-2019 |
| 5 | 15650044 | Kemi 1: The Lost Love of Freedom | 02-09-2019 | 09-09-2019 |

The format for making book loan data when formed will look like the table below:

**Table 2.** Example of Book Loan Format

| **Transaction** | **Text** | **Guide** | **Scientific** | **Novel** | **Encyclopedia** | **Biography** | **Short story** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

Information:

1. A novel is a written, narrative work of prose fiction; usually in the form of a story. Novelists are called novelists. The word novel comes from the Italian novella which means "a story, a piece of news".
2. Encyclopedia or encyclopedia, is a number of books that contain explanations of each branch of science arranged alphabetically or according to categories in a brief and concise manner.
3. A biography is a story or information about a person's life. A biography is more complex than just a list of dates of birth or death and data about a person's occupation, it also tells about the feelings involved in experiencing events. A biography written by the character himself is called an autobiography.
4. Scientific work is a research report, accompanied by, thesis, and so on.
5. Guidelines are also called guidebooks, for example books on raising chickens, oil palm gardening, tips for obtaining and tips for receiving scholarships abroad, and so on.
6. Scientific is a book that is arranged based on scientific principles. For example, books that are compiled based on research results and delivered in scientific language.
7. Texts are textbooks, textbooks, modules.

The book lending data that will be studied is for 1 month, namely September 2019 (Before the Covid-19 Pandemic) as many as 228 transactions and 7 book items. Transaction data processing will be carried out in several stages, namely as follows:

1. Before searching for patterns from borrowing data, look for all the names of the books contained in the transaction table to determine support per book item, this stage is looking for item combinations that meet minimum support. So the authors set a minimum 2% support rule.



**Figure 1.** Support List of Each Itemset

1. The final association rules are based on a minimum predetermined support and minimum confidence. The author determines confidence = 70% and support = 2%, it can be seen in the table below:



**Figure 2.** Association Rule Result Display

**4. Conclusion and recommendations**

*4.1. Conclusion*

From the descriptions in the chapters that have been discussed previously, conclusions can be drawn:

1. The choice of variables used greatly influences the resulting rule or knowledge.
2. The Apriori Algorithm with its association rules can provide effective information to describe the processes associated with library visitor patterns.
3. The resulting lending data pattern with a minimum support of 2%, a minimum of 70% confidence is a combination pattern of five items.
4. Students who borrow short stories have a 97% chance of borrowing a Biography.
	1. *Recommendations*

The development of this system is from a manual system or based on a knowledge base to a computerized system built on the basis of consideration of problems arising from the old system. The following are suggestions that can be shown for the future.

1. In this study, the authors tried one of the techniques used in determining the patterns of library visitors, namely the Apriori Algorithm. To get good pattern results, several or a combination of several prediction techniques can be used. Therefore it is necessary to carry out further research to compare the results of the patterns and determine which theory produces the best predictions.
2. To get a good visitor pattern, a complete data source is needed. Therefore, it is hoped that the library data storage process has been carried out electronically, facilitating the search for the data needed to carry out the knowledge discovery process such as determining the patterns of library visitors.

References

[1] Bakker E 2020 Implementation of Data Mining Clustering Library Data Using the K-Means Algorithm to Determine Additions to Book Collections in UPY Libraries *NATIONAL SEMINAR of Informatics Dynamics* (Yogyakarta: PGRI Yogyakarta University) pp 22–5

[2] Sastie DE, Suparni and Pohan AB 2020 Analysis of the Apriori Algorithm on Book Borrowing Patterns at the Ahmad Dahlan ITB Library *J. Media Inform. Budidarma* 4 136–43

[3] Aziz S, Sayuti A and Mustakim 2017 Application of the ARIMA Method for Forecasting Visitors to the UIN Suska Riau Library *National Seminar on Information, Communication and Industry Technology (SNTIKI)* (Pekanbaru: UIN Sultan Syarif Kasim Riau) pp 18-9

[4] Kurniawan E 2018 Implementation of Data Mining in Analysis of Patterns for Borrowing Books in Libraries Using the Association Rule Method *JURTEKSI (Jurnal Teknol. And Sist. Information)* 5 89–96

[5] Afdal M and Rosadi M 2019 Implementation of Association Rule Mining for the Analysis of the Layout of Books in the Library Using the Apriori Algorithm *J. Ilm. Engineering and Management. Sist. Inf.* 5 99–108

[6] Yanto R and Kesuma H In 2017 the Utilization of Data Mining for Placing Books in Libraries Using the Association Method *Jatisi* 4 1–10

[7] Amborowati A 2008 Designing and Making Data Warehouse at STMIK AMIKOM Yogyakarta Library *National Seminar on Science and Technology Application* (Yogyakarta) pp 39–52

[8] Santoso H 2017 Data Mining Preparation of West Lombok Regional Library Books Using the Apriori Algorithm *National Seminar on ICT and Social Sciences (SocioTech)* (Mataram: STMIK Bumigora) pp 25–35

[9] Budiyati E, Hurniningsih and Lusita MD 2020 Implementation of the Apriori Algorithm Method for Placing Books on the Library Shelf STMIK Jakarta STI & K *J. Inf. Syst. Informatics Comput.* 4 30–9

[10] Qomariyah N, Nafiiyah N and Hanifah AI 2018 Classification of Book Categories in the Informatics Engineering Library of Islamic University Lamongan JOUTICA 3 159–63

[11] Lingga D 2016 Application of the Apriori Algorithm in Predicting Book Supply at the SMA Dwi Tunggal Tanjung Morawa Library *Maj. Ilm. Inf. and Teknol. Ilm.* 11 18–22

[12] Wandi N, Hendrawan RA and Mukhlason A 2012 Development of a Book Search Recommendation System with Association Rule Excavation Using the Apriori Algorithm (Case Study of the Library and Archives of East Java Province) *J. Tech. ITS* 1 1–5

[13] Anas A 2016 Analysis of Apriori Algorithm to Obtain Book Borrowing Patterns at SMPN 3 Batanghari Library *J. Ilm. SISFO media* 10 628–41

[14] Anas A 2014 Analysis of Library Book Borrowing Patterns Using the Apriori Algorithm *J. Edik Inform.* 6 52–62

[15] Budhi GS, Handojo A and Wirawan CO 2009 Generalized Sequential Pattern Algorithm for Digging Sequential Data of Book Circulation at Petra UK Library *National Seminar on Information Technology Application (SNATI)* (Yogyakarta) pp 68–72