Utilization of data mining for placement of books in the library using the Apriori method

R R Waliyansyah*, M Novita and N D Saputro

Faculty of Engineering and Informatics, Universitas PGRI Semarang, Semarang, Indonesia

*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

A college library is a library located in a college or high school, academy and other higher education, which in essence is 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 teaching and learning activities 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 data on book lending transactions that are 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 algorithm can find data sets that appear most frequently in a data set [8]. By using a priori algorithm, recommendations will be given to librarians for book placement based on borrowed data [9], thus helping to improve services to the library [10].

The purpose of this research is to apply a priori algorithm to obtain set items of book borrowing with high frequency. Find associative rules by using a priori algorithm in getting book combinations to predict book inventory. Implementing a priori algorithm to predict book inventory with the help of Orange Biolab software.
2. Research materials and methods

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 tools (Tools Orange Biolab).

Apriori is an algorithm that is well known for searching frequent itemset using the association rule technique [11]. 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:

\[
\text{Support}(A) = \frac{\text{number of transactions containing } A}{\text{total transactions}} \quad (1)
\]

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

\[
\text{Support}(A, B) = P(A \cap B) \quad (2)
\]

\[
\text{Support}(A, B) = \frac{\sum \text{transactions containing } A \text{ and } B}{\sum \text{transactions}} \quad (3)
\]

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 \( A \rightarrow B \). The confidence value from the rule \( A \rightarrow B \) obtained the following formula:

\[
\text{Confidence} P(B|A) = \frac{\sum \text{transactions containing } A \text{ and } B}{\sum \text{transactions containing } A} \quad (4)
\]

Association rules are determined and selected should be sorted based on support x confidence. N rules are taken that have the greatest payoff. A rule that states the association between several attributes is often called an affinity analysis or market basket analysis [13]. Figure 1 shows association rule calculation stages.
3. Results and discussion

Several previous studies related to the Apriori Algorithm include the first is an Improvised a priori algorithm using frequent pattern tree for real time applications in data mining. Based on this algorithm, this paper indicates the limitation of the original Apriori algorithm of wasting time and space for scanning the whole database searching on the frequent itemsets, and present an improvement on Apriori [14]. The second is a research entitled Apriori-Based Algorithm for Dubai Road Accident Analysis. Studies conducted show that during weekend's peak accident time, majority of the drivers were intoxicated, while during week. On the basis of the results obtained, certain recommendations which aid in reducing the accident rate have been put forward [15].

In this study, apriori 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 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, the authors show that the data on book lending at the University of PGRI Semarang so far have not been well optimized, so that the data on borrowing is increasing every day. The data is only stored as campus archives or bookkeeping and there is no known benefit from the existing data. Example of data on book lending for September 2019 as seen in Table 1.

Table 1. Book borrowing data 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 Table 2 below:

Table 2. Example of a 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           |

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:

- Before searching for patterns from lending 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. The list of support for each itemset can be seen in Figure 2.

![Figure 2. Support list of each itemset.](image)

- The final association rules are based on a predetermined minimum support and minimum confidence. The author determines confidence = 70% and support = 2%, it can be seen in Figure 3 below:
4. Conclusion and recommendations

From the descriptions in the chapters that have been discussed previously, it can be concluded that the choice of variables used greatly affects the rule or knowledge produced. A priori algorithm with its association rules can provide effective information to describe the processes associated with library visitor patterns. The resulting borrowing data pattern with a minimum of 2% support, at least 70% confidence is a combination pattern of five items. Students who borrow short stories have a 97% chance of borrowing a Biography.

References

[1] Bakker E 2020 Implementasi Data Mining Clustering Data Perpustakaan Menggunakan Algoritma K-Means untuk Menentukan Penambahan Koleksi Buku di Perpustakaan UPY SEMINAR NASIONAL Dinamika Informatika (Yogyakarta: Universitas PGRI Yogyakarta) pp 22–5

[2] Sastie D E, Suparni and Pohan A B 2020 Analisa Algoritma Apriori Pada Pola Peminjaman Buku di Perpustakaan ITB Ahmad Dahlan J. Media Inform. Budidarma 4 136–43

[3] Aziz S, Sayut A and Mustakim 2017 Penerapan Metode ARIMA untuk Peramalan Pengunjung Perpustakaan UIN Suska Riau Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) (Pekanbaru: UIN Sultan Syarif Kasim Riau) pp 18–9

[4] Kurniawan E 2018 IMPLEMENTASI DATA MINING DALAM ANALISA POLA PEMINJAMAN BUKU DI PERPUSTAKAAN MENGGUNAKAN METODE ASSOCIATION RULE JURTEKSI (Jurnal Teknol. dan Sist. Infomasi) 5 89–96

[5] Afdal M and Rosadi M 2019 PENERAPAN ASSOCIATION RULE MINING UNTUK ANALISIS PENEMPATAN TATA LETAK BUKU DI PERPUSTAKAAN MENGGUNAKAN ALGORITMA APRIORI J. Ilm. Rekayasa dan Manaj. Sist. Inf. 5 99–108

[6] Yanto R and Kesuma H Di 2017 Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Jatisi 4 1–10

[7] Amborowati A 2008 PERANCANGAN DAN PEMBUATAN DATA WAREHOUSE PADA PERPUSTAKAAN STMK AMIKOM YOGYAKARTA Seminar Nasional Aplikasi Sains dan Teknologi (Yogyakarta) pp 39–52

[8] Santoso H 2017 Data Mining Penyusunan Buku Perpustakaan Daerah Lombok Barat Menggunakan Algoritma Apriori Seminar Nasional TIK dan Ilmu Sosial (SocioTech)
(Mataram: STMIK Bumigora) pp 25–35

[9] Budiyati E, Hurniningtih and Lusita M D 2020 IMPLEMENTASI METODE ALGORITMA APRIORI UNTUK PENEMPATAN BUKU PADA RAK PERPUSTAKAAN STMIK JAKARTA STI&K J. Inf. Syst. Informatics Comput. 4 30–9

[10] Qomariyah N, Nafiiyah N and Hanifah A I 2018 KLASIFIKASI KATEGORI BUKU PADA PERPUSTAKAAN TEKNIK INFORMATIKA UNIVERSITAS ISLAM LAMONGAN JOUTICA 3 159–63

[11] Lingga D 2016 PENERAPAN ALGORITMA APRIORI DALAM MEMPREKISI PERSEDIAN BUKU PADA PERPUSTAKAAN SMA DWI TUNGGAL TANJUNG MORAWA Maj. IIm. Inf. dan Teknol. IIm. 11 18–22

[12] Wandi N, Hendrawan R A and Mukhlason A 2012 Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalian Association Rule Menggunakan Algoritma Apriori (Studi Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur) J. Tek. ITS 1 1–5

[13] Anas A 2014 Analisa Pola Peminjaman Buku Perpustakaan Menggunakan Algoritma Apriori J. Edik Inform. 6 52–62

[14] Bhandari A, Gupta A and Das D 2015 Improvised apriori algorithm using frequent pattern tree for real time applications in data mining Procedia Comput. Sci. 46 644–51

[15] John M and Shaiba H 2019 Apriori-Based Algorithm for Dubai Road Accident Analysis Procedia Comput. Sci. 163 218–27