

Book Borrowing Patterns Using Apriori Algorithm in East Java Library and Archives

Ellen Maria Padwasih¹, Aries Dwi Indriyanti²

^{1,2} *State University of Surabaya, Surabaya, Indonesia*

ellen.19050@mhs.unesa.ac.id, ariesdwi@unesa.ac.id

ABSTRACT

This study analyzes book borrowing patterns at the East Java Provincial Library and Archives Service using the Apriori algorithm. Data from borrowing transactions between January 16 and March 10, 2023, were analyzed to identify relationships between book titles. This data mining study involves data preprocessing, checking for missing values and duplicates, as well as transforming data into a binary matrix. The analysis results show that item combinations with a minimum support value of 1.2% can be identified. Association rules such as AC, AE → AG have a support value of 0.012 and a confidence level of 40.9%. Further analysis reveals more complex patterns, such as AE, AJ → AC, with a confidence level of 57.1%. This research contributes to understanding book borrowing behavior and offers solutions to improve library management efficiency, such as book recommendation systems and more appropriate collection arrangement strategies. Thus, libraries can develop services that are more responsive to community needs.

Keyword: Apriori Algorithm, Borrowing Pattern Analysis, Data Mining, Association, Analysis.

Article Info:

Article history:

Received January 24, 2025

Revised July 08, 2025

Accepted October 13, 2025

Corresponding Author

Ellen Maria Padwasih

State University of Surabaya, Surabaya, Indonesia

ellen.19050@mhs.unesa.ac.id

1. INTRODUCTION

Libraries play a vital role in society by providing access to a wide range of information and books. Therefore, libraries must deliver excellent services that meet the needs of the community [1]. The East Java Provincial Library and Archives Service is responsible for managing libraries across East Java. In 2023, the number of visitors to regional libraries in East Java reached 172,873, highlighting the significant role libraries play in serving the community [2]. This figure underscores the importance of libraries as centers for literacy and information for the public. However, with the increasing number of visitors and the diversity of their needs, managing library collections and services has become more complex.

A more modern and adaptive approach to library management is required to meet the literacy needs of the public in the digital age [3]. As information needs become increasingly complex, the application of data mining technologies like the Apriori Algorithm is crucial for effectively identifying book borrowing patterns. Understanding these patterns not only supports

more relevant collection management but also enhances library services by offering book recommendations tailored to visitors' preferences.

Book borrowing patterns are a key aspect of library management [4]. Data mining uses association rules to reveal complex patterns and relationships between dataset items through a two-stage process: analyzing item associations and consumer behavior, and identifying frequently co-occurring items [5]. The most appropriate method for analyzing these patterns is the Apriori Algorithm, a critical tool in data mining used to discover association patterns and calculate itemset frequencies [6]. The Apriori Algorithm was utilized to uncover patterns and relationships between factors using dataset-based analysis [7]. By identifying popular item combinations and extracting meaningful association rules, libraries can leverage the Apriori Algorithm to gain deeper insights into reader preferences and provide more personalized services [8]. This study employs the Apriori method to analyze book borrowing data at the East Java Library, aiming to identify associations between books frequently borrowed together.

The Apriori Algorithm holds great potential for improving library services. First, it is effective in identifying borrowing patterns, enabling libraries to understand visitor preferences and manage collections optimally. Second, through its association analysis capabilities, the algorithm can provide relevant book recommendations to visitors, enhancing user experience and facilitating the discovery of suitable books. Thus, libraries can improve efficiency, better meet visitors' needs, and raise awareness of the importance of library services [9].

This research benefits the East Java Provincial Library and Archives Service by optimizing resource management, enhancing operational efficiency, and improving visitor experiences. Furthermore, it contributes to developing methods for analyzing borrowing patterns that can be applied by other libraries. Through this innovative approach, the study aims to make a positive impact not only locally but also as a significant reference for modern library management in other regions.

The study is relevant and useful for the East Java Provincial Library and Archives Service in its efforts to enhance library services. Additionally, the findings are expected to contribute to the development of methods for analyzing book borrowing patterns that can be adopted by other libraries. Therefore, this research is crucial in supporting the development of modern libraries that are responsive to community needs in the digital era. The study is also expected to benefit other researchers interested in advancing data mining research and provide insights for libraries looking to implement data mining techniques in managing book borrowing data.

2. METHODS

Knowledge Discovery in Databases (KDD) is a process that helps us find useful patterns and relationships in large amounts of data. It's a systematic approach to identifying important and meaningful patterns from complex data sets. Data Mining is the key part of KDD that helps us analyze and predict things from data. The ultimate goal of KDD is to turn data into useful knowledge. KDD is a powerful tool that helps organizations stay up-to-date with what their customers want, how they behave, and what they do [10]. This study employs the KDD approach. The application of KDD in libraries aims to analyze book borrowing patterns to uncover relationships between book titles. The research workflow is as follows:

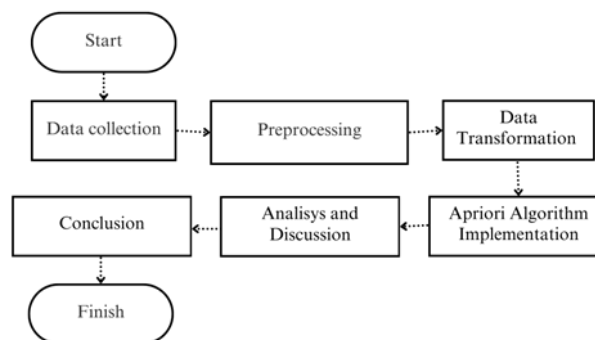


Figure 1. Research Flowchart

3. RESULTS AND DISCUSSION

3.1 Data Collection and Selection of Book Borrowers Dataset

The dataset includes 1,955 registered borrowers, reflecting the reading interest and library usage activities of the community from various age groups and backgrounds. The book borrowing data in the library is well-organized, with key attributes such as borrowing serial number, bibliographic data (title, author, publication year, edition), and book classification based on the library's system. This data reflects the public's reading preferences and is classified into various categories of knowledge. Before being analyzed using the Apriori algorithm, the data undergoes a selection process, which includes eliminating duplicates, handling missing values, and refining the data format.

Table 1. Dataset Structure

No	Deadline	Reg-number	Bibliographic Data	Klass Number	Member's Number	Member's Name	Gender	Group of Age
1	1/16/2023 0:00	00675/DPK/2019	Jurassic world: claws and teeth poster book	500 - Ilmu Murni	22121900007	ACHM AD DZAKI ARRAFI F	M	Child (6 - 11)
2	1/16/2023 0:00	12.080/BPK/P/2015	Aku jago menggambar Princes / Irwan Nuswantoro	700 - Kesenian dan Olahraga	22121900007	ACHM AD DZAKI ARRAFI F	M	Child (6 - 11)
3	1/30/2023 0:00	03387/DPK/P/2022	Ayah dan sirkus pohon : original story / Andrea Hirata ; penyunting, Dhewiberta, Nurani Nura	800 Kesusasteraan	22061600005	AGHEL IA TAUFA NI HIDIAN TI	F	Last Teens (17 - 25)

3.2 Preprocessing Data

One of the key steps in preprocessing is the removal of columns that are considered irrelevant to the analysis objectives. In this study, columns such as Due Date, Member ID, Klass Number, Gender, and Age Group were removed as they do not directly contribute to identifying book borrowing patterns.

Instead, the study focuses on columns deemed most essential, namely Member Number and Bibliographic Data. Member Number is used as a unique identifier for each borrower, while

Bibliographic Data reflects information about the books being borrowed, which is central to analyzing the relationships between items in the dataset.

Table 2. Result of Preprocessing

Member's Number	Bibliographic Data
22121900007	Jurassic world : claws and teeth poster book
22121900007	Aku jago menggambar Princes / Irwan Nuswantoro
22061600005	Ayah dan sirkus pohon : original story / Andrea Hirata ; penyunting, Dhewiberta, Nurani Nura

3.3 Data Transformation

The transformation of bibliographic data facilitates analysis by converting it into structured new columns. This process involves the following steps:

Transformation Steps:

- 1) Encoding: Converting book titles into short labels (letters or symbols).
- 2) Binary Representation: Changing the values of the columns into binary format (1 for borrowed, 0 for not borrowed).

The result is a structured binary matrix, which simplifies the analysis of book borrowing patterns and helps identify books that are frequently borrowed together.

Table 3. Result of Encoding Data

Code	Bibliographic Data
A	Jurassic world : claws and teeth poster book
B	Aku jago menggambar Princes / Irwan Nuswantoro
C	Ayah dan sirkus pohon : original story / Andrea Hirata ; penyunting, Dhewiberta, Nurani Nura

The next step in data preprocessing is checking for missing values in the dataset. The preprocessing process for book borrowing data involves several key steps:

Data Cleaning

- 1) Duplicate Check and Removal: Identifying and removing duplicate data to avoid redundancy.
- 2) Missing Value Handling: Checking for and addressing missing values.

Data Organization

- 1) Grouping Data by Member Number: Organizing data according to the unique Member Number.
- 2) Converting Data into Transaction Format: Structuring the data into an organized transaction format.

Table 4. Result of Dataset Transformation

Member's Number	A	B	C	D	.	.	.	FO

22121900007	1	1	0	0	.	.	.	.
22061600005	0	0	1	0	.	.	.	.
35780800001	0	0	0	1	.	.	.	.
35780800005	0	1	0	0	.	.	.	.
20010200008	0	1	0	0	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

This analysis can help identify borrowing patterns, provide book recommendations, and improve library collection management

3.4 Algorithm Implementation

After obtaining the frequent itemsets, the next step is to build association rules using the Create Association Rules operator. At this stage, association rules are generated based on key parameters such as support and confidence. This transformation is crucial because the Apriori algorithm requires data in binary format (1/0 or TRUE/FALSE) to indicate the presence or absence of item.

For a 1 item, the support value is calculated using the formula:

$$\text{Support (A)} = X/Y$$

which:

A: item type.

X: Number of transactions containing item A.

Y: Total number of transactions in the database.

For two items or item combinations, the support value is calculated with the following formula:

$$\text{Support (A,B)} = P(A \cap B) = \frac{Z}{Y}$$

which:

B: Book Classification item

A,B: Transactions involving item combinations A and B (book classifications).

$P(A \cap B)$: Probability of transactions containing both items A and B.

Z: Number of transactions containing both items A and B.

After high-frequency patterns are identified, the second stage involves forming association rules that describe the relationships between items. These rules ($A \rightarrow B$) indicate the likelihood that item B will appear after item A in a transaction, with the Confidence value calculated using the following formula:

$$\text{Confidence (A} \rightarrow \text{B)} = P(A \cap B) = \frac{Z}{Y}$$

which:

$A \rightarrow B$ The rule showing the relationship between item A and item B.

$P(B|A)$: The probability of item B being included when item A is already present in a transaction.

Z: The number of transactions containing both items A and B.

X: The number of transactions containing item A.

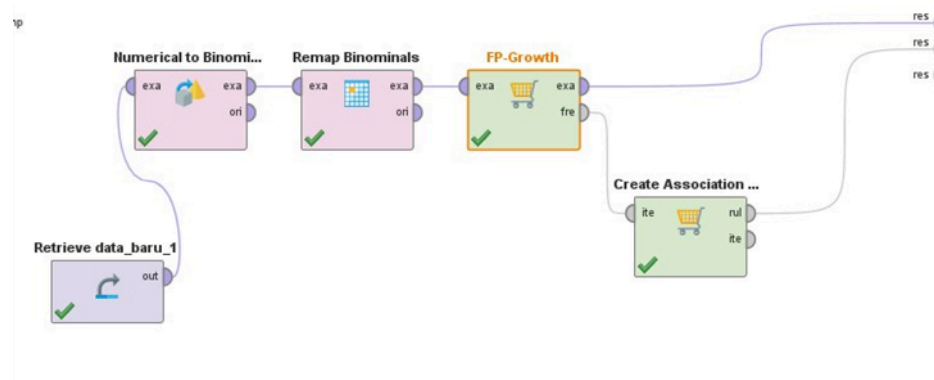


Figure 2. Implementation of Algorithm

In this study, the minimum support value was set at 2%. This value is used as a threshold to filter items to be used in forming association patterns. The number of transactions containing an item will be compared to the total transactions.

Items meeting the criteria will proceed to the next stage, which involves forming more complex itemset combinations (frequent 2-itemsets, 3-itemsets, and so on). This process will be carried out iteratively until combinations that frequently occur together are identify.

Table 5. Result of Associations Pattern

No	Premises	Conclusion	Support	Confidence
1	AC, AE	AG	0.012	0.409
2	AE, AG	AC	0.012	0.409

Based on the displayed association rules, two rules indicate relationships between items in the dataset, derived from the Frequent 1-itemset schema:

The First Rule $AC, AE \rightarrow AG$, this rule shows that if items AC and AE appear together, there is a 40.9% probability that item AG will also appear, with a support value of 0.012, representing approximately 1.2% of the total transactions.

The Second Rule: $AE, AG \rightarrow AC$, this rule indicates that if items AE and AG appear together, there is a 40.9% probability that item AC will also appear, with the same support value of 0.012.

The Confidence value of 40.9% is significant, suggesting a strong association pattern among these items.

The analysis of Frequent 2-itemsets focuses on combinations of two items that frequently appear together in the dataset. The results reveal complex patterns with strong and specific relationships among the items.

Table 6. Result of Frequent 2 itemset

Premises	Conclusion	Support	Confidence
AE, AJ	AC	0.005263	0.444444
AG, AR	AC	0.005263	0.444444
AG, AM	AD	0.005263	0.444444
AD, AR	AI	0.006579	0.454545
AG, AH	AC	0.006579	0.5

AF, AR	AD	0.005263	0.5
AG, AS	AI	0.005263	0.5
AI, AS	AG	0.005263	0.5
AD, AN	AE	0.006579	0.555556
AI, AR	AD	0.006579	0.555556
AG, AJ	AC	0.005263	0.571429

The study using the Apriori algorithm reveals significant association patterns among items in the book borrowing transaction data. The results have important implications for library management, including grouping books based on borrowing patterns, developing book recommendation systems, and optimizing library layouts. For instance, books AE, AJ, and AC exhibit a strong relationship with a Confidence of 57.1%. These findings can enhance user convenience, reduce search time, and improve user experience through relevant recommendations.

CONCLUSION

This study successfully implemented the Apriori algorithm to analyze significant association patterns in library book borrowing transaction data. The results revealed combinations of books frequently borrowed together, such as AE, AJ, and AC, with a Confidence of 57.1%. These findings indicate a strong relationship between these books and have the potential to improve library management efficiency.

The research outcomes are valuable for the development of more sophisticated and efficient library management systems. Grouping books based on borrowing patterns and implementing book recommendation systems can enhance user accessibility and improve the overall user experience. This study opens opportunities for developing better systems and supporting operational and strategic decision-making in library management.

Based on these findings, the following recommendations can be applied to enhance library management:

1. **Optimizing Library Layout**
Placing frequently borrowed books on adjacent shelves can simplify user access and reduce search time.
2. **Developing Advanced Analytics**
Using other algorithms, such as clustering, to analyze user preferences in more detail can improve recommendation systems and library services.
3. **Improving Data Quality**
4. **Ensuring that borrowing transaction data is accurate and consistently updated will enhance the effectiveness of analysis and the relevance of recommendations.**

ACKNOWLEDGEMENTS

The author wishes to extend sincere gratitude to the East Java Provincial Library and Archives Service for their invaluable support and assistance in the completion of this research. Appreciation is also expressed for the opportunity to access and utilize the necessary data for this research.

REFERENCES

- [1] H. F. A. Hermawati, "Layanan Perpustakaan", Yogyakarta: Penerbit Andi, 2013.
- [2] W. Pratama, "Pengunjung Perpustakaan Daerah di Jatim Mencapai 173.874", *suarasurabaya.net*, 14 September 2023, [online]. Tersedia: <https://www.suarasurabaya.net/kelanakota/2023/pengunjung-perpustakaan-daerah-di-jatim-mencapai-172-873/>. [Diakses; 10 Oktober 2023]
- [3] M. Ramadhan dan A. Anjasmara, "Manajemen Perpustakaan dalam Meningkatkan Layanan Perpustakaan," *Jurnal Intelek dan Cendekiawan Nusantara*, vol. 1, no. 4, 2024, pp. 5619-5625.
- [4] A. Nurseptaji, A. Arey, F. Andini, dan Y. Ramadhani, "Implementasi Metode Waterfall Pada Perancangan Sistem Informasi Perpustakaan," *Jurnal Dialektika Informatika (Detika)*, vol. 1, no. 2, 2024, pp. 49-57.
- [5] S. Kumar et al., "Data Analytics Pandemic Management using MapReduce and Apriori Algorithm," *Procedia Computer Science*, vol. 29, pp. 1279-1288, 2023.
- [6] M. Arifin, "Implementasi Data Mining Pada Prediksi Pemesanan Menggunakan Algoritma Apriori (Studi Kasus: Kimia Farma)," *Pelita Informatika: Informasi dan Informatika*, vol. 8, no. 3, 2020, pp. 353-356.
- [7] Hassan, Md., et al., "An Apriori Algorithm-Based Association Rule Analysis to detect Human Suicidal Behaviour," *Procedia Computer Science*, vol. 219, pp. 1279-1288, 2022, doi: 10.1016/j.procs.2023.01.412.
- [8] X. Zhang dan J. Zhang, "Analysis and research on library user behavior based on apriori algorithm," *Measurement: Sensors*, vol. 27, pp. 100802, 2023.
- [9] D. E. Satie et al., "Analisa Algoritma Apriori Pada Pola Peminjaman Buku di Perpustakaan ITB Ahmad Dahlan," *Jurnal Media Informatika Budidarma*, vol. 4, no. 1, 2020, pp. 136.
- [10] M. Kumar, et al., "Knowledge Discovery and Intelligent Data Mining," 2022, doi: 10.36647/MLAIDA/2022.12. B1. Ch012.
- [11] A. Cevoli dan E. Esposito, "From pool to profile: Social consequences of algorithmic prediction in insurance," *Big Data & Society*, vol. 7, no. 2, pp. 2053951720939228, 2020.
- [12] I. Sugiyarto, R. Irawan, dan D. Rosiyadi, "Pengelompokan Dampak Gempa Bumi Dan Kerusakan Pada Wilayah Berpotensi Gempa Di Provinsi Sumatra Barat: Indonesia," *Journal of Students 'Research in Computer Science*, vol. 2, no. 2, pp. 211-222, 2021.
- [13] S. Sudriyanto, F. Listrianti, dan J. Jamal, "Implementasi Algoritma C4.5 Untuk Memprediksi Kesesuaian Gaya Belajar Siswa Sekolah Dasar," *COREAI: Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi*, vol. 3, no. 2, 2022.
- [14] P. Suryati, "Analisis Pola Peminjaman Buku dengan Menggunakan Algoritma Apriori," *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, no. 1, pp. 17-23, 2022.
- [15] Y. Khoiri dan T. Ma'sum, "Manajemen Perpustakaan dalam Menumbuhkan Minat Baca," *JURNAL PIKIR: Jurnal Studi Pendidikan dan Hukum Islam*, vol. 9, no. 1, pp. 15-34, 2023.