Library Book Modeling Data Using the Association Rule Method with Apriori Algorithm in determining Book Placement and Analysis of Book Loans

Dedy Rahman Prehanto, Aries Dwi Indriyanti, Ginanjar Setya Permadi, Tanhella Zein Vitadiar, F D Jayanti

📄 PDF

Abstract

The library is one place that is able to provide information for the means of teaching and learning activities. The existence of daily operational activities in the library can increase the amount of data borrowing books. The more data, if not processed it becomes useless. Therefore, a data modeling was made to process book borrowing data. The purpose of this study was to implement the association rule method with a priori algorithm for modeling data on book borrowing at the As'ari Hasyim University library. From the application of this method, it will be used to recommend book placement and analyze book borrowing. This study produced an application that was used to determine associative rules between one book and another. From the resulting modeling, the calculation process is then carried out so that it can be known to borrow the same book every day. The results of this study percentage of support value and confidence value from the results of predictions with the association rule method with a priori algorithm on 600 data of borrowing transactions support values obtained an average of 6%, while for confidence values obtained on average of 67% - 100%.

International Journal of Advanced Science and...

Not yet assigned quartile

SJR 2020
0

powered by scimagojr.com

Make a Submission

ELSEVIER

Indexed by
Scopus

Downloads

Paper Format

Copyright Form

Library Book Modeling Data Using the Association Rule Method with Apriori Algorithm in determining Book Placement and Analysis of Book Loans

Dedy Rahman Prehanto,
Informatics Engineering Department, Universitas Negeri Surabaya, Indonesia
Aries Dwi Indriyanti,
Informatics Engineering Department, Universitas Negeri Surabaya, Indonesia
Ginanjar Setya Permadi,
Departement of Information Management, Faculty of Information Technology, Hasyim Asy'ari University, Jombang - Indonesia
Tanhella Zein Vitadiar,
Departement of Information Management, Faculty of Information Technology, Hasyim Asy'ari University, Jombang - Indonesia
F D Jayanti
Departement of Informatics, Faculty of Information Technology, Hasyim Asy'ari University, Jombang - Indonesia

Abstract

The library is one place that is able to provide information for the means of teaching and learning activities. The existence of daily operational activities in the library can increase the amount of data borrowing books. The more data, if not processed it becomes useless. Therefore, a data modeling was made to process book borrowing data. The purpose of this study was to implement the association rule method with a priori algorithm for modeling data on book borrowing at the Asy'ari Hasyim University library. From the application of this method, it will be used to recommend book placement and analyze book borrowing. This study produced an application that was used to determine associative rules between one book and another. From the resulting modeling, the calculation process is then carried out so that it can be known to borrow the same book every day. The results of this study percentage of support value and confidence value from the results of predictions with the association rule method with a priori algorithm on 600 data of borrowing transactions support values obtained an average of 6%, while for confidence values obtained an average of 67% - 100%.

Keywords: *Library, modeling, association rule, apriori algorithm*

1. Introduction

Library as a means of supporting teaching and learning activities that provide many book references to increase knowledge. Similarly, the library of Hasyim Asy'ari University presents a place for services to students in learning activities of the academic community. The existence of daily operational activities in the library can increase the amount of large book loan data. Based on the amount of data will increasingly increase if it is not used will become trash and not useful, only for the sake of archival documentation only. Therefore, the data can be used as input material in a data modeling to process book loan data which can later provide useful information (Anas, 2016).

Data modeling using the association rule method using a priori algorithm can simplify the calculation of book loan transactions made a pattern related to data mining (Kurniawan et al., 2014). Association rule as one of the ways of data mining that describes the transactions that often arise in an association group. With the modeling of book lending data using the association rule, students can find out the location of the books borrowed at the same time. So that students have no difficulty in finding books to borrow, because students already know the location of books arranged by the library (Kurniawan et al., 2014). Use of a priori algorithm is used to help explore the pattern of associations in book lending transactions that occur every day and to find out which books are often borrowed. So the

results of this modeling produce recommendations for determining the layout of books and the addition of the number of books relating to lending transactions.

Data modeling and data processing into useful information using the association rule method and a priori algorithm as the calculation processing is presented in the form of an information system. Information systems are a way of presenting information that can provide added value. Added value can be obtained in the form of information based on real data that is processed in such a way as to produce something useful for the recipient. One example of its use is planning, initiating, organizing, controlling activities and presenting information based on data processing. Resulting in useful information as a reference for determining the final decision. The simple understanding of information systems is that there must be input, process and output.

Modeling and processing of lending transaction data at the Hasyim Asy'ari University library as a recommendation for determining the layout and increasing the number of books presented in the form of a website. So as to facilitate system users in modeling and data processing. Thus the information system on determining the layout and number of procurement of books can be put to good use.

2. Method
2.1 Information Systems

The simple understanding of information systems is that there must be input, process and output. Information systems are a way of presenting information that can provide added value. Added value can be obtained in the form of information based on real data that is processed in such a way as to produce something useful for the recipient. One example of the use of information systems as planning, initiating, organizing, controlling activities and presenting information based on data processing. Resulting in information in the form of recommendations that are useful as consideration for final decision making.

2.2 Association Rule

Association rule is a data mining technique for finding association rules between a combination of items. The association rule for borrowing books in a library is a place to know how likely it is for a customer to borrow a computer book together with a mathematics book (Kusrini and Lutfi, 2009). Based on this knowledge library staff can arrange the placement of goods or design marketing campaigns using discount coupons for certain book combinations.

Association analysis is also often referred to as market basket analysis. Association analysis is also known as one of the data mining techniques that is the basis of various other data mining techniques. Especially one of the stages of association analysis called high frequency pattern analysis attracts the attention of many researchers to produce efficient algorithms (Kusrini and Luthfi, 2009). This concept itself is derived from the terminology of market basket analysis, namely the search for relationships of several products in a purchase transaction. This technique looks for possible combinations that often appear (frequent) from an itemset (Witten dkk., 2005).

Association rule is an expression of implications in the form of $X \rightarrow Y$, where X and Y are a disjoint itemset $(X \cap Y) = \emptyset$. In the association rule, we can calculate support and confidence. Confidence states how often the items in Y appear in transactions that contain X (Hermawati, 2013). In transactions containing item X there may be items Y also in them, denoted $X \rightarrow Y$, where X and Y are disjoint itemsset, notated $X \cap Y$, the collection of these transactions is called an itemset denoted by IK (k = 1 , 2, ... m). If there is an itemset that has as many items as K, then it is called a k-itemset. This association rule will produce a rule that uses how big the relationship between X and Y, and two measures are needed for this rule, namely support and confidence (Defit, 2013).
Support is the possibility that X and Y appear together notated:

$$Support = \sum \frac{borrowed\ items\ at\ once}{total\ number\ of\ transactions} \times 100\% \qquad (1)$$

Whereas confidence is the possibility of Y appearing when X also appears, denoted:

$$Confidence = \sum \frac{borrowed\ items\ at\ once}{number\ of\ transactions\ in\ the\ antecedent\ section} \times 100\% \qquad (2)$$

The first step in the association rule is to produce all possible itemset with the possibility that the itemset that appears with m-items is $2 \wedge m$. Because of the amount of computation to calculate frequent itemset, which compares each itemset candidate with each transaction, there are several approaches to reduce the computation, one of which is a priori algorithm (Defit, 2013).

Association rules want to provide this information in the form of an "if-then" or "if-then" relationship. This rule is calculated from probabilistic data. Association analysis is also known as a data mining method which is the basis of various other data mining methods (Wirdasari dan Calam, 2011).

The importance of an associative rule can be determined by two parameters, support (supporting value), which is the percentage of combination of the item. In the database and confidence (certainty value), namely the strong relationship between items in the associative rules. Association analysis defined a process for finding all associative rules that met the minimum requirements for support and the minimum conditions for confidence (Wirdasari dan Calam, 2011).

There are several algorithms that have been developed regarding the rules of association, but there is one classic algorithm that is often used is the a priori algorithm. The basic idea of this algorithm is to develop frequent itemset. By using one item and recursively developing a frequent itemset with two items, three items and so on to frequent itemset of all sizes (Wirdasari dan Calam, 2011).

To develop frequent sets with two items, you can use frequent set items. The reason is if the set of one item does not exceed the minimum support. In general, developing sets with fc-items uses frequent sets with k-1 items that were developed in the previous step. Each step requires one check throughout the contents of the database (Wirdasari dan Calam, 2011).

The importance of an associative rule can be determined by two parameters, support (supporting value), which is the percentage of combination of the item. In the database and confidence (certainty value), namely the strong relationship between items in the associative rules. Association analysis defined a process for finding all associative rules that met the minimum requirements for support (minimum support) and the minimum requirements for confidence (Wirdasari dan Calam, 2011).

2.3 Apriori

A priori algorithm is an algorithm known by R. Agrawal and R. Srikant in 1994. A priori algorithm aims to find the frequent itemset that is run on a set of data. In the k-iteration all itemsset that have k-items will be found, called k-itemset. The main characteristic of a priori algorithm is that all subsets of frequent itemsets also apply members of frequent itemsets (Virgiawan and Mukhlas, 2013). Its use is to produce association rules, using the "if-then" pattern. A priori algorithm uses an iterative approach known as level wise search, where k-groups of products are used to explore (k + 1) -groups of products or (k + 1) -itemset (Anas, 2016).

A priori algorithms use knowledge about frequent known itemset, to process further information. For high frequency patterns are patterns of items in a database that have a frequency or support above a certain threshold called the minimum support term. This high frequency pattern is used to develop associative rules and also some other data mining techniques (Anas, 2016). A priori algorithm is included in the type of association rules in data mining. Apart from a priori, this group includes the Generalized Rule Induction method and the Hash Based Algorithm. Rules that state the association between several attributes are often called anffinity analysis or market basket analysis (Santoso, 2007).

Association analysis or association rule is a data mining technique for finding associative rules between a combination of items. A priori algorithm which aims to find frequent itemsets is run on a set of data. A priori analysis defines a process to find a priori rules that meet the minimum requirements for support and minimum requirements for confidence (Santoso, 2007). The principle of a priori algorithm is to collect the number of single items, get large items, get candidate pairs, count → large pairs of items, get candidate triplets, count → large triplets of items and so on, as a clue: each subset of a frequent itemset must be frequent (Anas, 2016).

The two main processes in the a priori algorithm are the steps that will be taken to get frequent itemset. Although the a priori algorithm is easier to understand and implement compared to other algorithms that are actually applied to the association rule process (Santoso, 2007). A priori algorithm has a disadvantage that is to do frequent itemset searches, the a priori algorithm must scan the database repeatedly for each combination of items. This causes the amount of time needed to scan the database.

In addition it takes a large generate candidate to get a combination of items from the database (Anas, 2016).

A priori algorithm is used to find frequent itemset that meets minimum support then get a rule that meets the minimum confidence of frequent itemset earlier. This algorithm controls the development of itemset candidates from frequent itemset results with support based pruning to eliminate unattractive itemset by setting minimum support. The principle of this a priori is when an itemset is classified as a frequent itemset, and vice versa (Wandi et al., 2012). The way this algorithm works generates new candidates for the k-itemset from the frequent itemset in the previous step and calculates the k-itemset's support value. Items that have a support value below the minimum support will be deleted. The algorithm stops when no more frequent new itemset is generated (Wandi dkk., 2012).

A priori algorithm is used to find association rules that meet the support and confidence limits. How it works a priori is to analyze a collection of goods purchased together on several transactions. Calculation of a collection of data patterns that appear in a database through several iterations or repetitions. Iteration i counts all the i data sets (data sets containing element i) that occur frequently. Each iteration consists of two steps, namely candidate generation (candidate determination) and candidate counting and selection (selection and calculation of candidates) (Haryanto et al., 2011). This a priori algorithm describes a basic algorithm proposed by Agrawal and Srikant in 1994 to find frequent itemset for Booelan association rules. A priori algorithms include types of association rules in data mining. Association analysis or association rule is a data mining technique for finding the rules of a combination of items (Kusrini dan Luthfi, 2009).

One of the stages of association analysis that attracts many researchers to produce efficient algorithms is the analysis of high frequency patterns. The importance of an association can be known by two benchmarks, namely: support and confidence. Support (supporting value) is the percentage of the item combination in the database, while confidence (certainty value) is the strong relationship between items in the association rules (Kusrini dan Luthfi, 2009).

3. Results and Discussion

Transaction data obtained from historical data in a year are calculated in transactions and borrowed items. This Transaction Data is used to find out loan transactions. Transaction data from historical data numbers can be summarized at Table 1.

Table 1 Transaction Data

Transaction	Borrowed item	Transaction	Borrowed item
1	30 Menit Mahir Membuat Jaringan Komputer	2	Metode Penelitian Praktis
1	Koleksi Lengkap Software Hacking	2	Psikologi Pendidikan
1	*Kriptografi Untuk Keamanan Jaringan*	3	Rekayasa Perangkat Lunak
1	Panduan Lengkap Microsoft Windows Server 2008	3	Pengantar Studi Islam
1	Jaringan Komputer	3	Studi Islam Kontemporer
1	Evaluasi Pembelajaran	3	Psikologi Pendidikan
1	Jaringan Komputer Linux	3	Dasar-Dasar Ilmu Pendidikan
1	Membangun Sistem Jaringan *Wirelees*	3	Modern Power System Control
1	Analisis Sistem Infromasi	3	Kecerdasan Tiruan
1	Metodelogi Penelitian	3	Pengantar Jaringan Komputer

2	Pengantar Filsafat Islam	3	Dasar-dasar Ilmu Perpustakaan
2	Pemrograman Bahasa Assembly	3	Teori Dan Modul Praktikum Jaringan
2	Tip & Trik Registry Windows 7	4	Agama Einstein
5	Ilmu Ushul Fiqih	4	Mikro Ekonomi, Teori Pengantar
5	Aplikasi Logika Fuzzy Untuk Pendukung Keputusan	4	Studi Islam Asia Tenggara / peny.: Zaenuddin Fananie, M. Thayibi
5	Pengantar Ilmu Tarekat Kajian Historis Tentang Mistik	4	Perencanaan Bahasa Pada Era Globalisasi
5	Epistomologi Tafsir Kontemporer	4	Pengantar Statistika
5	Pengantar Ekonomi	4	Pokok-Pokok Materi Statistik 1
5	Islam: A Short History	4	Sistem Digital
5	Marketing	4	Akhlaq Tasawuf 1
5	Prinsip dan Teknik Evaluasi Penganggaran	4	Sistem Pendukung Keputusan
5	Manajemen Bank Syari'ah	4	Sistem Kendali Dasar
		5	Microsoft Visual C# 2010

After the transaction data table is formed then determine the tabulation table. Tabulation tables are used to calculate the number of loans for each item. From frequent itemset we can get the same amount of book borrowing when one transaction is calculated in a combination between data.

After the combination is formed, then determine the analysis results obtained from the combination results calculated with support and confidence, then the results of the support and confidence results with the calculations obtained in the following formula:

$$Support = \sum \frac{borrowed\ items\ at\ once}{total\ number\ of\ transactions} \times 100\% \qquad (3)$$

$$Confidence = \sum \frac{borrowed\ items\ at\ once}{number\ of\ transactions\ in\ the\ antecedent\ section} \times 100\% \qquad (4)$$

Graph analysis results from the calculation process obtained can be displayed in graphical form that illustrates the value of the possibility to borrow as a comparison as shown in Figure 4.
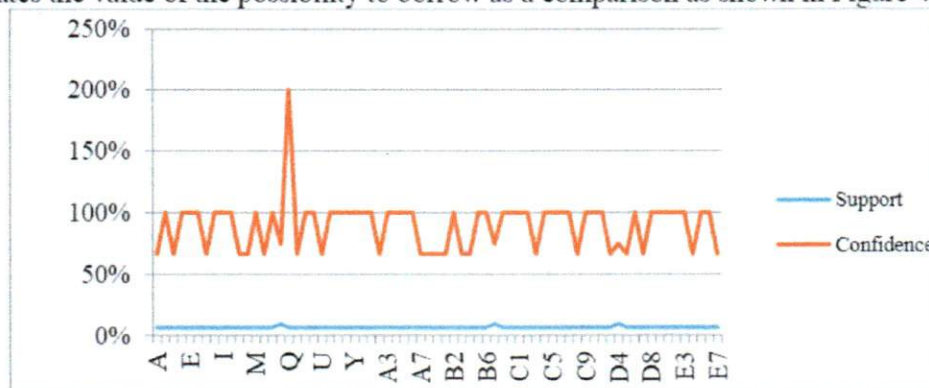

Figure 4 Grafik Hasil Prediksi

Then the same item is made a combination to make a rule involving the value of support and value of confidence. After that, calculate the support value and confidence value of the rules that have

been selected from the combination. Then the results of the calculation of support and confidence will be used to determine the rules that will be taken from the value of the confidence that has been determined as a value that will be categorized as the selected rule. The final results of calculations using a priori algorithm can be seen in Table 2.

Table 2 The results of the analysis process with the highest confidence value

Rule	*Support*	*Confidence*	*Support X Confidence*
If borrowed Jaringan Komputer then borrow Metodologi Penelitian Pendidikan dan Sosial (Kuantitatif dan Kualitatif)	6%	67%	0.043010753
If borrowed Reabilitas dan Validitas then borrow Metodologi Penelitian Pendidikan dan Sosial (Kuantitatif dan Kualitatif)	6%	100%	0.064516129
If borrowed Jaringan Komputer then borrow Pengantar Studi Islam	6%	67%	0.043010753
If borrowed Analisis Sistem Infromasi then borrow Dasar-Dasar Ilmu Pendidikan	6%	100%	0.064516129
If borrowed Pisikolog Remaja then borrow Dasar-Dasar Ilmu Pendidikan	6%	100%	0.064516129
If borrowed Membangun Sistem Jaringan Wirelees then borrow Dasar-Dasar Ilmu Pendidikan	6%	100%	0.064516129
If borrowed Pengantar Jaringan Komputer then borrow Dasar-Dasar Ilmu Pendidikan	6%	67%	0.043010753

4. Conclusion

Based on the research results of book lending data modeling using the association rule method with the a priori algorithm, conclusions and suggestions are obtained as a percentage of the value of support and confidence values from the results of the prediction using the association rule method with a priori algorithm to 600 lending transaction data, the value of support is obtained an average of 6% while for confidence values obtained on average 67% -100%.

5. References

[1] Lathifah , E. L., Suparman. 2018. Design of Mathematics Student Worksheet based on Discovery Learning Approach to Improving Mathematical Representation Ability Students of Grade VII Junior High School in Indonesia. International Journal of Engineering & Technology. 7 hal 26-30.
[2] Limbong, T., Simarmata, J., dkk. 2018. The implementation of computer based instruction model on Gost Algorithm Cryptography Learning. 2nd Nommensen International Conference on Technology and Engineering hal 420.
[3] Yusri, R., Nurmi, N., Delyana, H. 2019. Development of ICT integrated project based learning student worksheet. International Conference on Mathematics and Science Education 1157.
[4] Satchwell, R. E., Loepp, F. L. 2002. Designing and Implementing an Integrated Mathematics, Science, and Technology Curriculum for the Middle School. Journal of Industrial Teacher Education Volume 39, Number 3.
[5] Fadila, A., Dasari, A., Setiyaningsih, Septiana, R., Rosyid, A. 2019. The Development of Electronic Flash Worksheet Based on Adobe Flash Cs6 on Fraction Numbers in the Seventh Grade of Junior High School. IOP Conf. Series: Journal of Physics: Conf. Series 1155.
[6] Permadi, G.S., Kusworo, A., and Gernowo, R. 2018. Application Mail Tracking Using RSA Algorithm As Security Data and HOT-Fit a Model for Evaluation System E3S Web of Conferences vol 31, Semarang.

[7] Anggraini. R., dkk. 2018. Electronic Module Design with Scientifically Character Charged Approach on Kinematics Material Learning to Improve Holistic Competence of High School Students in 10th Grade. IOP Conf. Series: Materials Science and Engineering 335.

[8] Wibawa, S. C., Wahyuningsih, Y., dkk. 2018. Online test application development using framework CodeIgniter. The Consortium of Asia-Pacific Education Universities (CAPEU) hal 296.

[9] Dewi, W. S., Festiyed. F., Sari, S. Y. 2018. The Study of Literacy Reinforcement of Science Teachers in Implementing 2013 Curriculum. IOP Conf. Series: Materials Science and Engineering 335.

[10] Rahayu, H., Yani, A. P. 2017. Pengembangan Lks Berdasarkan Eksplorasi Tanaman Obat Suku Pekal Pada Materi Keanekaragaman Hayati Jurnal Pendidikan dan Pembelajaran Biologi 1(1): 37-45.

[11] Maharani, L., Rahayu, D. I., dkk. 2019. Toondoo Application Based on Contextual Approach: Development of Comic Learning Media. YSSTEE Journal of Physics: Conf. Series 1155.

[12] Zaher, A., Damaj, I. W. 2018. Extending STEM Education to Engineering Programs at the Undergraduate College Level. iJEP – Vol. 8, No. 3.

[13] Vitadiar, T. Z., Farikhin, Surarso, B. 2018. Production Planning and Planting Pattern Scheduling Information System for Horticulture. E3S Web of Conferences vol 31, Semarang.

[14] Mashuri, C., Suryono, Suseno, J. E. 2018. Prediction of Safety Stock Using Fuzzy Time Series (FTS) and Technology of Radio Frequency Identification (RFID) for Stock Control at Vendor Managed Inventory (VMI). E3S Web of Conferences vol 31, Semarang.

[15] Suryono, S., Suseno, J. E, Mashuri, C., Sabila, A. D., Nugraha, J. A. M., Primasiwi, M. H. 2018. RFID Sensor for Automated Prediction of Reorder Point (ROP) Values in a Vendor Management Inventory (VMI) System Using Fuzzy Time Series. E3S Web of Conferences vol 31, Semarang.