Search Of Favorite Books As A Visitor Recommendation of The Fmipa Library Using CT-Pro Algorithm

Sufiatul Maryana[*], Lita Karlitasari[*]

[*]*Pakuan University, Bogor, Indonesia*
Corresponding Author: sufiatul.maryana@unpak.ac.id

Abstract. The library of Faculty of Mathematics and Natural Science (FMIPA) has a collection of books and other print media, total of 2,678 books with 7237 visitors and 2148 borrowers. The available book search system was very helpful for visitors to find the required books. Especially if the system has features recommended of books. In the provision of book recommendations used one of the data mining techniques, namely association rule mining techniques or excavation of association rules. In the development of this recommendation system, KDD (Knowledge Discovery from Database) model was used. The data used was the transaction history of borrowing book with the category of "chemistry", for the last 5 (five) months, that is September 2014 - February 2015. The excavation technique of this association rule has 2 (two) main process, they are: frequent patterns and rules. To find frequent patterns, a CT-PRO algorithm was used. The minimum value of support used was 1 and 2. Once the pattern is found, the confidence value of each pattern was calculated. The minimum value of confidence used ranges from 10% to 100%. The recommendation rule was based on calculating the value of this confidence. The comparison of minimum support values indicates that the greater value of minimum support then the less borrowing pattern was generated, and vice versa. The comparison of minimum confidence value shows that the greater of minimum confidence value then the less recommended rule given..

Keywords: Library, Recommended System, Knowledge Discovery from Database (KDD), Association Rule Mining, CT-PRO Algorithm

---

I. INTRODUCTION

Pakuan University as one of the best educational institutions in Bogor city currently has the mission of becoming a superior educational institution, independent and characterize. Where each unit in it work together to be able to realize the mission of the Pakuan University. One of unit that requires an accurate and fast information system is in the library section.

Currently, many other forms of libraries are developed, called as an electronic library that used computers and Internet as the media. The current FMIPA library already has a Library Information System (SIPUS) which has been operated since September 2014. There are also 2678 books, consisting of various disciplines such as Biology, Chemistry, Mathematics, Computers, and Pharmacy. Meanwhile, the borrowing transactions that begin to operate through SIPUS from September 2014 to 31 December 2015 was 2148 transactions. With a total of 7237 visitors.

For visitors that searching for references from a topic, then the book tracking system already available in this library was the main target. Searching of references usually required more than one book, then the facility of book recommendation system will greatly help the visitor.

The use of CT-PRO algorithm in searching for frequent patterns of borrowing transactions (any book that borrowed in a borrowing transaction) is expected to provide book recommendations on the library tracking system of the Pakuan University Faculty of MIPA (Maryati *et al.*, 2015). Thus, the library can meet the needs of visitors in search of certain books with a sufficient number of copies based on the recommendations of the book's search.

II. RESEARCH METHODS

The methodology used in collecting data to support the research are:

1. Research

In this study the data was taken from field research and literature (Library Research).

a. Field research is the collection of data directly in the location of research by observation or study the files that exist.

b. Library research (Library research) is a method of collecting data by studying and understanding the theories and various literature related to the topic discussed.

2. Interview

This method is carried out to obtain an overview of the visitor and the overall borrowing transaction that relates to the research title as the report material.

3. Literature

Collection of materials that are related to the discussion in the study.

Research Stages

The research used the model of Knowledge Discovery from Database (Bell and Jason, 2015.), (Harrington and Peter, 2012). This model has 7 (seven) phases, namely: data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. The Knowledge Discovery from Database (KDD) model cycle as shown in Figure 1.
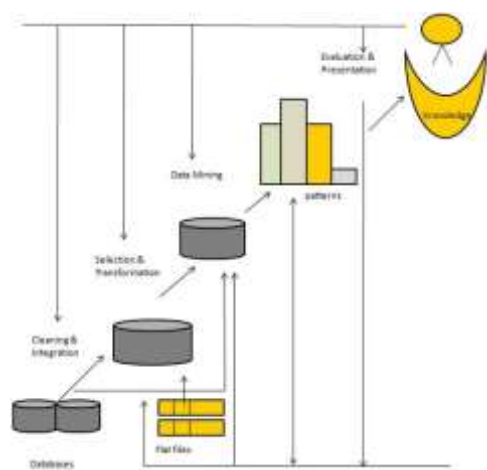

Figure 1. KDD Model Cycle

1. Data Cleaning

At this stage, deletions are performed for the noise data and irrelevant (Brown and Meta 2014). This stage is very important, because the result of data mining process depends on the quality of the selected data. The process in the data cleaning stage are: delete duplicate records or records that contain lots of missing value.

2. Data Integration

This stage is done to unify data from various data storage locations (database or data warehouse).

3. Data Selection

At this stage was the process data selection of a databinding data. The selected data related to further data mining process. The unneeded data can be removed.

4. Data Transformation

At this stage was to perform the data format adjustments for data mining, such as changing the format of some data or the overall data, to fit the tools during the process of data mining calculation. After this stage, the data was ready to be processed.

5. Data Mining

At this stage, the selection of data mining techniques that match to the purpose was done (Ahlemeyer-Stubbe *et al.*, 2014). After that, the technique was applied to previously prepared data. The results of this stage are the patterns of interest data.

6. Pattern Evaluation

At this stage, testing was done to the data patterns that found in the data mining process. Because, not all patterns are really "interesting" based on the measure. If it meets the standard measure, then the pattern used as knowledge.

7. Knowledge Presentation

At this stage, visual techniques was designed to present the knowledge of data mining process to be easily understood by users and also to implement the design into the technology base (Han *et al.*, 2012).

III. RESULTS AND DISCUSSION

Determine of Sample Data

The number of borrowing transactions from September 2014 to December 2015 was 2148 transactions. To do the calculation analysis, sample data were taken as many as 25 transactions.

Table 1. Sample Data

No.	ID Buku	Judul Buku
1	1512	Kimia Organik
2	1512	Kimia Organik
3	1516	FESSENDEN FESSENDEN Kimia Organik
	1476	Analisa Kimia Kuantitatif
4	1512	Kimia Organik
	1722	Analisis Anorganik Kualitatif Makro dan Semikro
5	1512	Kimia Organik
6	1699	Kimia Universitas Asas & Struktur E5
	1516	FESSENDEN FESSENDEN Kimia Organik
7	1699	Kimia Universitas Asas & Struktur E5
	1516	FESSENDEN FESSENDEN Kimia Organik
8	1478	Konsep Dasar Kimia Analitik
	1476	Analisa Kimia Kuantitatif
9	1478	Konsep Dasar Kimia Analitik
	1476	Analisa Kimia Kuantitatif
10	1722	Analisis Anorganik Kualitatif Makro dan Semikro
11	1476	Analisa Kimia Kuantitatif
	1699	Kimia Universitas Asas & Struktur E5
12	1699	Kimia Universitas Asas & Struktur E5
	1478	Konsep Dasar Kimia Analitik
13	1476	Analisa Kimia Kuantitatif
	1699	Kimia Universitas Asas & Struktur E5
15	1699	Kimia Universitas Asas & Struktur E5
	1478	Konsep Dasar Kimia Analitik
16	1722	Analisis Anorganik Kualitatif Makro dan Semikro
	1715	Kimia Anorganik Dasar
17	1722	Analisis Anorganik Kualitatif Makro dan Semikro
	1715	Kimia Anorganik Dasar
18	1722	Analisis Anorganik Kualitatif Makro dan Semikro
	1730	Kimia Lingkungan
19	1722	Analisis Anorganik Kualitatif Makro dan Semikro
	1730	Kimia Lingkungan
20	1512	Kimia Organik
	1602	Atlas Berwarna & Teks Biokimia

21	1512	Kimia Organik
	1516	FESSENDEN FESSENDEN Kimia Organik
22	1512	Kimia Organik
	1552	Fessenden & Fessenden Kimia Organik E3
23	1512	Kimia Organik
24	1722	Analisis Anorganik Kualitatif Makro dan Semikro
	1516	FESSENDEN FESSENDEN Kimia Organik
25	1699	Kimia Universitas Asas & Struktur E5

To facilitate the calculation, then in table 1 will be simplified as shown in table 2.

Table 2. Simplification of sample data

No	Item	No	Item
1	1512	14	1699, 1478
2	1512	15	1722, 1715
3	1476, 1516	16	1722, 1715
4	1512, 1722	17	1722, 1730
5	1512	18	1722, 1730
6	1699, 1516	19	1512, 1722
7	1699, 1516	20	1512, 1602
8	1476, 1478	21	1512, 1516
9	1476, 1478	22	1512, 1552
10	1722	23	1512, 1722
11	1699, 1476	24	1516
12	1699, 1478	25	1699
13	1699, 1476		

For the minimum value of support taken from the distribution of borrowing transaction data as shown in Table 3:

Table 3. Distribution of borrowing transaction data

Jumlah Buku per Transaksi	Jumlah Transaksi	Persentase
1	113	62%
2	42	38%

So the minimum value of support ranges from 1 - 2. For this calculation analysis used minimum value support (min_sup) of 2.

Specifies the Frequent Itemset with CT-PRO Algorithm

Specify the frequent itemset was done to find the frequent patterns, a CT-PRO algorithm was used. CT-PRO algorithm useful for finding books that are often borrowed simultaneously (Sucahyo *et al*., 2004), (Gupta, 2011). The steps of CT-PRO algorithm as below:
1. Calculate the frequency of all transactions item.
2. Remove the item value of below min_sup. Items 1552 and 1602 are eliminated because of their value < min_sup.
3. Enter the item into the global table header. Compiled from the largest frequency item to the smallest.
4. Sort items in each transaction by index
5. Perform a search of frequent itemset locally from the last index, index 8.

The results of frequent patterns of all local search as shown in Table 4:

Table 4. Result of frequent patterns with CT-PRO algorithm

no	item	frequency
1	1730, 1722	2
2	1715, 1722	2
3	1478, 1699	2
4	1478, 1476	2
5	1516, 1699	2
6	1476, 1699	2
7	1722, 1512	3

The flowchart of CT-PRO algorithm as shown in Figure 2 and 3.

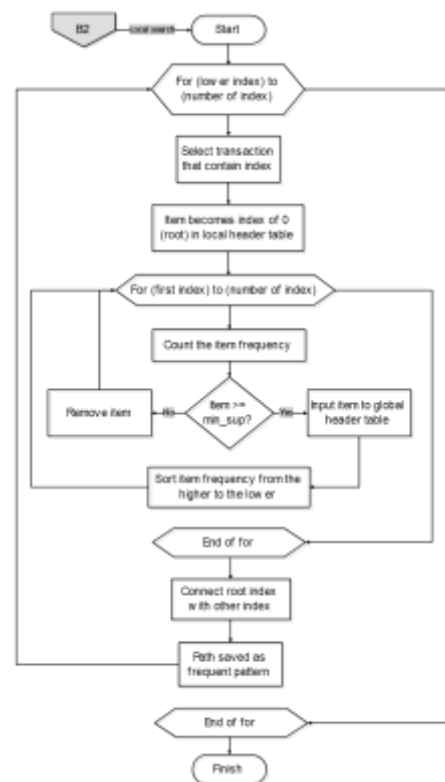Figure 2. Flowchart of CT-PRO algorithm

Figure 3. CT-PRO algorithm flow diagram (continued)

Result

The system interface page in Figure 4 shows the description of:

1. Search page

At this page, user can inputs a keyword in the search box of a book title.

Figure 4 Search page interface

2. Search Results Page

On this page, the search results displayed from the keywords entered by the user as in Figure 5. To view book details, the user can clicks the **see details** button.

Fugure 5 Search results page

3. Search Details page

On this page, search of book details displayed as in Figure 6, 7 and 8. Resulted the data mining calculations and recommended books.

Figure 6. Display of search detail page (book description)

Figure 7. Detailed search page (calculation result)

Figure 8. Display of search detail page (book recommendation)

Discussion

In association rule mining techniques, there are 2 (two) main processes (Wandi *et al*., 2012), namely:

a. Searching for frequent patterns (the pattern of items that often appear)

b. Rule determination

The process of frequent patterns search was used to search for books that are often borrowed simultaneously, using the CT-PRO algorithm. The minimum value of support used was 1 and 2. Minimum value of support was obtained from the distribution of borrowing transaction data.

The comparison of minimum value of supports shown in Table 5 and Figure 9:

Table 5. Comparison of total items that meet the minimum support

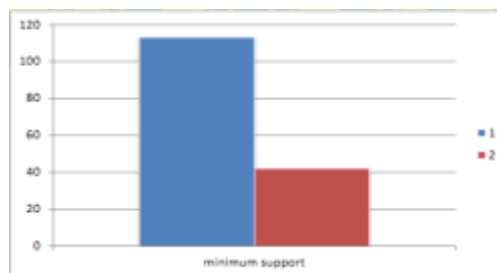| Minimum Support	Jumlah Item	Persentase
1	113	62%
2	42	38%

Figure 9. Chart comparison of minimum support

The table and the minimum support comparison chart indicate that the greater the minimum value of support, the less the item will be generated.

For the determining process of the rules, used the formula of confidence value. The value of confidence used to determine how high a trend of an item appears along with other items. The minimum value of confidence used ranges from 10% - 100%. The determination of minimum confidence value based on research conducted by Wandi, Hendrawan and Mukhlason (2012) with the same case study (see References). The result comparison of confidence values to the generated rules shown in Table 6 and Figure 10:

Table 6. Comparison of total rules based on minimum confidence value

Minimum Confidence	Jumlah *rules*
10%	13
20%	13
30%	10
30%	7
40%	5
50%	3
60%	3
70%	3
80%	3
90%	3
100%	3

Figure 10. Chart Comparison of Minimum Confidence

From the table and the minimum confidence chart showed that the greater the minimum value of confidence, the less rules produced, and vice versa. It means that the less rules generated, then the less book recommendations can be given and vice versa.

IV. CONCLUSION

This study used association rule mining technique and CT-PRO algorithm to give recommendation of books at FMIPA-UNPAK Library. The required data was the history of borrowing transaction and books list. Prior to the process of data mining, these data must be prepared to produce accurate calculation results. For this purpose, the KDD (Knowledge Discovery from Data) model was used, which consists of: data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation.

Associating rule mining techniques have 2 (two) main processes, namely: the search pattern of frequent items (frequent patterns) and the determination of rules. To find the frequent patterns, a CT-PRO algorithm was used. CT-PRO algorithm useful for finding books that are often borrowed simultaneously. To select books that are often borrowed, a minimum value of support was used. Once the pattern was found, the confidence value of each pattern was calculated. The confidence value used to measure the likelihood of borrowed book along with other books. A pattern that meets the minimum confidence value, becomes a rule. These rules will be used as the basis for recommending a book.

The comparison of minimum support values indicates that the greater the minimum value of support, the less the borrowing book pattern was generated, and vice versa. In comparison the minimum value of confidence shows that the greater the minimum confidence value, the less recommended rules are given, and vice versa.

REFERENCES

Ahlemeyer-Stubbe, Andrea. Coleman, Shirley. 2014. *A Practical Guide to Data Mining for Business and Industry*. WILEY, United Kingdom.

Bell, Jason. 2015. *Machine Learning: Hands-On for Developers and Technical Professionals*. John Wiley & Sons, Inc, Indiana Polis, Canada.

Brown, Meta S. 2014. *Data Mining For Dummies*. John Wiley & Sons, Inc, New Jersey, Canada.

Dwi Maryati, Sri Setyaningsih, Lita Karlitasari. *Penerapan metode association rule mining untuk rekomendasi penelusuran buku dengan menggunakan algoritma CT-PRO*. Skripsi, FMIPA, Universitas Pakuan. 2015

Gupta, Bharat. 2011. FP-Tree Based Algorithms Analysis: FPGrowth, COFI-Tree and CT-PRO. International Journal on Computer Science and Engineering (IJCSE). 3 : 2691 – 2699.

Han, Jiawei. Kamber, Micheline. Pei, Jian. 2012. *Data Mining Concepts and Techniques 3rd Edition*. Morgan Kaufmann, USA.

Harrington, Peter. 2012. *Machine Learning in Action*. Manning, Shelter Island, New York.

Sucahyo, Yudho Giri. Gopalan, Raj. P. 2004. CT-PRO: A Bottom-Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure.

Wandi, Nugroho. Hendrawan, Rully A. Mukhlason, Ahmad. 2012. Pengembangan Sistem Rekomendasi Penelusuran Buku dg Penggalian Association Rule Menggunakan Algoritma Apriori. Jurnal Teknik Institut Teknologi Sepuluh November (ITS). 1 : A445 – A449.