

# HASIL CEK\_Zahrotun, Jones\_Library, Association Rule, FP-Growth, study program.

*by Lisna Zahrotun, Anna Hendri Soleliza Jones Fp-growth Algorithm For  
Searching*

---

**Submission date:** 20-Sep-2022 01:01PM (UTC+0700)

**Submission ID:** 1904328551

**File name:** orrowing\_Transactions\_PAtterns\_And\_Study\_Program\_Suitability.pdf (661.81K)

**Word count:** 2546

**Character count:** 13351



## Fp-Growth Algorithm For Searching Book Borrowing Transaction Patterns And Study Program Suitability

Lisna Zahrotun<sup>1</sup>, Anna Hendri Soleliza Jones<sup>2</sup>

Department of Informatic, Faculty of Industrial Engineering, Universitas Ahmad Dahlan  
Ringroad Selatan, Kampus 4 UAD, Yogyakarta 55191

e-mail: lisna.zahrotun@tif.uad.ac.id<sup>1\*</sup>, annahendri@tif.uad.ac.id<sup>2</sup>

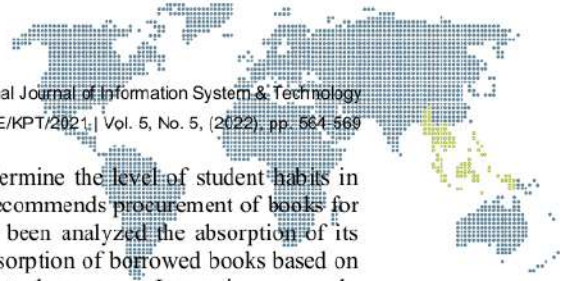
### Abstrak

The current development of data has reached a sizeable amount. This is due to the development of the world of information technology which consists of data in it. One technique that can handle abundant data is data mining. Data mining methods are widely used to perform large amounts of data analysis. In the academic field, analysis can be used to determine the patterns of students and lecturers. Whereas in library transactions, analysis can be carried out to determine the patterns of existing book borrowing. This is done to determine the tendency of students with certain study programs to borrow any uku transactions. In this study, the aim of this research is to analyze the patterns of borrowing books from the Ahmad Dahlan University library, which includes borrowing transaction data and the book owner's study program. In addition, in this study, a percentage analysis of the suitability of the borrower study program and the book owner's study program was also carried out. The stages in this research include data collection, data cleaning, data selection, data transformation, searching for association patterns using the FP-Growth method and pattern evaluation. The test used in this research is the lift ratio. The results of this study are publications in international journals that are in the draft process. Apart from that, the results of this study provide information on the analysis of patterns of lending books in libraries using the FP-Growth method. The resulting pattern is 103 patterns with a support count value of 5 and a confident 10% with the 2 itemset rule, this means that the level of book borrowing is still low. While the results of the analysis of the suitability of books in the study program with the borrower were 31% in accordance with the study program, namely Pharmacy and Public Health Sciences, meaning that there were 69% of students who borrowed books from the library that were not in accordance with their study program.

**Keywords:** Library, Association Rule, FP-Growth, study program.

### 1. Introduction

The library is a collection of information that can be accessed by the public to find references or borrow materials / information. The information provided by the library can be physical or digital. One of the functions of the library is a place for borrowing books which is a provider of information and knowledge. The quality of the library can be seen from how well the book lending and borrowing transaction function is carried out. Based on library interviews, a university in Yogyakarta also has more than 24,397 collections of book titles totaling 67,453 copies. Transactions that occur in the library every day reach 100 transactions, all of which have been recorded in the system. However, the large number of books means that there are also many books that are not actively borrowed by students. Therefore, it takes a step to find book borrowing patterns to determine the level of student habits in borrowing books. Apart from that every days reached 100 transactions which have all been recorded in the system. However, the large number of books means that there are also many books that are not actively borrowed by students. Therefore, it



takes a step to find book borrowing patterns to determine the level of student habits in borrowing books. In addition, each study program recommends procurement of books for each year. However, this available book has never been analyzed the absorption of its borrowing. Therefore it is important to know the absorption of borrowed books based on the borrower's study program and the book owner's study program. In previous research, we have conducted research related to data mining, namely searching for patterns of alumni data associations [1], drug data patterns [2], grouping customer data [3]. Apart from data mining, research on text mining is also carried out [4][5][6].

4 One way to solve the problems in the library is by implementing the Association rule. association rule is a data mining technique used to find associative rules between a combination of items. Frequent Pattern Growth (FP-Growth) is an a priori algorithm that can be used to determine the most frequent itemsets in a set of data. FP-Growth is faster than the Apriori algorithm [7]. By using the FP-Tree concept, the FP Growth algorithm also has a good ability to look for association patterns [8]. Several studies have been conducted to find out patterns of lending library books. As research conducted by [9] and on other studies conducted [10]. Both studies have been carried out and both are still using the a priori algorithm, which is basically the process of calculating the frequent itemset is still relatively slow, so this study has determined the eclat algorithm so that the frequent itemset calculation process is faster. In addition, research conducted at the library has not implemented the suitability of study program books with borrower study programs and still uses book IDs for pattern search calculations.

## 2. Research Methodology

### 2.1 Data Mining

Data mining is a process of finding important knowledge from large collections of data that have been stored in databases [11]. Such knowledge cannot be found in simple ways. Data mining uses certain techniques (methods) to find knowledge including classification, clustering and linkage analysis. According to [12] Data mining tasks can be grouped into two categories namely:

- a) Predictive mining is concerned with classification methods, regression and deviation detection.
- b) Descriptive mining will derive and investigate patterns (correlation, trend, cluster) of important relationships from data sets. The main method of descriptive mining is clustering, association rules and sequential mining.

### 2.2. FP-Growth

6 FP-Growth is an improvement on the a priori algorithm. This algorithm is used to define the most frequent itemset in a data set [13]. The FP-Growth algorithm uses the concept of a development tree in searching for frequent itemsets. The characteristic of the FP-Growth algorithm is that the data structure used is a tree called the FP-Tree. By using the FP-Tree, the FP-Growth algorithm can directly extract itemset from the FP-Tree. Excavation of frequent itemset using the FP-Growth algorithm will be carried out by generating a data structure tree or called the FP-Tree [14][15].

### 2.3. Factor of Interest

Support is an occurrence rule for all transactions in the dataset [11]. Minimum support is a value that must be met by the rules:

$$\text{Support}(A \rightarrow B) = P(A \cup B) = \frac{\sum A \cup B}{\sum \text{Transaction}} \quad (1)$$

### 2.4. Confidence

Confidence is the opportunity for an item to appear together with other items to appear. If it is formed with notation in  $A \rightarrow B$ , then trust is defined as how often B occurs





when A also occurs. Minimum Confidence value is defined as the minimum to be fulfilled by the rules.

$$Confidence(A \rightarrow B) = P(A|B) = \frac{P(A \cup B)}{P(A)} = \frac{\text{support count}(A \cup B)}{\text{support count}(A)} \quad (2)$$

### 2.5. Lift Rasio

Lift Ratio is used to find out the correlation between the items in the rule. If the value of the lift rule is  $> 1$  then Positively correlated, if the value of the lift rule  $< 1$  then negatively correlated, if the value of the lift rule = 1 then it's independent (not combined).

$$Lift(A \rightarrow B) = P(A|B) = \frac{P(A \cup B)}{P(A) \cdot P(B)} = \frac{\text{confidence}(A \cup B)}{\text{support}(A)} \quad (3)$$

### 2.6. Frekuent Pattern Growth

FP-Growth is one of the algorithms used to solve the Association Rule case. This algorithm has two steps. The first two steps, compression is performed on the database based on frequently occurring items to create a Frequent Pattern Tree (FP-Tree). Second, separation is performed on the database results into a compressed database condition. FP-Tree is a special feature that distinguishes the FP-Growth algorithm from the a priori algorithm. FP-Tree has two features, first starting from the root which is named "null". Then from the root to form a sub-tree consisting of certain items. Second, each node contains three important information, namely the item label (indicates the type of item (item ID) it is represented by the node), support (the number shows the number of transaction paths through that node or also known as frequency), connecting pointers (link nodes) as a liaison. Between vertices and items in each path between cells, marked with a line connecting the pointer arrows.

## 3. Result and Discussion

The results of this study are the association pattern and the percentage of suitability of book borrowers with the book owner's study program at a university library. Before the pattern search is carried out, the first step taken is taking the data of the library book lending transaction that will be processed. The data that has been successfully displayed is then carried out by a process of selecting and transforming the data. The results of the selection and data transformation then carried out a pattern search using the FP Growth method. After finding several patterns, the next step the borrower's suitability pattern and graph with the book owner's program of study. An image of the research is to check the correlation of the patterns produced using the lift ratio. And the final step is to display the results of steps is shown in Figure 1.

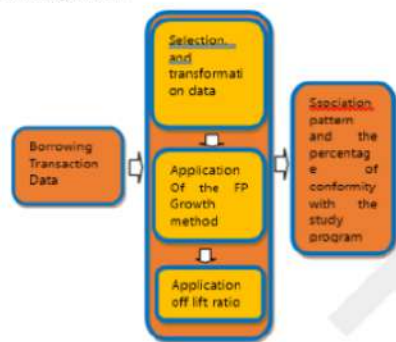


Figure 1. Research Diagram



### 3.1. Implementation

The first step is data retrieval which is the initial process by entering an excel file in the form of initial book loan transaction data. The transaction data table can be seen in Figure 2.

Uploaded Data

	0	1	2	3
6	6	0748V20182003	074/B/V/2018/DA/3/C.7	Buku Ajar Kesehatan Iu...
7	9	1838II2017LPSI	083/B/I/2017/LPSI/3/C.L.	Tuntunan Dashed Paakti...
8	8	1218V2016FAR	1216/B/V/2016/FAR/3/C.L.	Taksonomi Unsu Dasar G...
9	9	1388V2007DK	138/B/V/2007/DA/3	Buku Saku Hematologi
10	10	0428II2017FAR	042/B/I/2017/FAR/3/C.9	Buku Ajar Vogel Kinia ...
11	11	0528V111996FAR	052/H/V/11/1996/FAR/3	Biologi Jilid 2
12	12	3048V2006FAR	304/B/V/2006/DA/3	Berlaku Organisasi I...
13	13	7888V12005DK	788/B/V/12/2005/DA/3	Manajemen Pecuasaan d...
14	14	2218K1002TK	221/B/K/2002/TK/3	Introduction to Chemi...
15	15	0988II2017LPSI	098/B/I/2017/LPSI/3/C.L.	Kuliah Simul Media
16	16	0378V2018DM3	037/B/V/2018/DM/3/C.2	CITIZEN 4.0 : Masa Jarak...

Figure 2. Load Data

The next stage, namely data selection is carried out before the information mining stage. The selected data is stored in the file. at this stage remove unused variables. The data that has been successfully selected can be seen in Figure 3.

In [15]: #f	0	1	2	3	4	5	6	7	8	9	10
0	1	100802017FAR	1008BA/2017FAR/C.14	Kita Fermani Angas	100802020K	Aula Helmi Puri	20	-	Name	Kategori	Fermani
1	2	4048V2017DK	4048V/2017DK/C.4	Metode Penelitian Kuantitatif Kualitatif dan RAO	173804004	WISLITJAU FALZAH	20	-	Name	Kategori	Manajemen Pendidikan SI
2	3	3888V2014DK	3888V/2014DK	Etna	100802000	NURMILA FEBRIANA	20	-	Name	Kategori	Ilmu Kesehatan Masyarakat
3	4	3888V2014DK	3888V/2014DK	Etna	100802000	NURMILA FEBRIANA	20	-	Name	Kategori	Ilmu Kesehatan Masyarakat
4	5	1308V2018DK	1308V/2018DK/C.5	KEMASAP Lingsingan Smp K3	100804007	Lili Febiana Putriana	20	-	Name	Kategori	Ilmu Kesehatan Masyarakat

Figure 3. Selected Data

The next stage, namely this transformation is carried out by changing the data variable into valid or ready for the data mining process. The results can be seen in Figure 4

Data Transformation

	Inventaris	Prodi
0	Farmasi	Farmasi
1	Ilmu Kesehatan Masyarakat	Manajemen Pendidikan SI
2	Ilmu Kesehatan Masyarakat	Ilmu Kesehatan Masyarakat
3	Ilmu Kesehatan Masyarakat	Ilmu Kesehatan Masyarakat
4	Ilmu Kesehatan Masyarakat	Pasca Sarjana Psikologi Seire
5	Ilmu Kesehatan Masyarakat	Pasca Sarjana Psikologi Seire
6	lembaga Pengembangan Studi Ilm	Biologi
7	Farmasi	Biologi
8	Ilmu Kesehatan Masyarakat	Biologi
9	Farmasi	Teknik Kimia - SI
10	Farmasi	Psico Biologi

Figure 4. Data Transformation

The final step is Result FP-growth to display the results of the association rules where there is an input form to input minimum confidence and minimum support with the minimum confidence 0.7 and minimum support conditions 8. The results of the FP-growth result can be seen in Figure 5.



```
D: [18] |> library(fpGrowth)
D: [19] |> patterns = fpGrowth_kbn_frequent_patterns(cek_buku_issues(), 5)
D: [20] |> patterns
D: [21] |>
D: [22] |>
D: [23] |>
D: [24] |>
D: [25] |>
D: [26] |>
D: [27] |>
D: [28] |>
D: [29] |>
D: [30] |>
D: [31] |>
D: [32] |>
D: [33] |>
D: [34] |>
D: [35] |>
D: [36] |>
D: [37] |>
D: [38] |>
D: [39] |>
D: [40] |>
D: [41] |>
D: [42] |>
D: [43] |>
D: [44] |>
D: [45] |>
D: [46] |>
D: [47] |>
D: [48] |>
D: [49] |>
D: [50] |>
D: [51] |>
D: [52] |>
D: [53] |>
D: [54] |>
D: [55] |>
D: [56] |>
D: [57] |>
D: [58] |>
D: [59] |>
D: [60] |>
D: [61] |>
D: [62] |>
D: [63] |>
D: [64] |>
D: [65] |>
D: [66] |>
D: [67] |>
D: [68] |>
D: [69] |>
D: [70] |>
D: [71] |>
D: [72] |>
D: [73] |>
D: [74] |>
D: [75] |>
D: [76] |>
D: [77] |>
D: [78] |>
D: [79] |>
D: [80] |>
D: [81] |>
D: [82] |>
D: [83] |>
D: [84] |>
D: [85] |>
D: [86] |>
D: [87] |>
D: [88] |>
D: [89] |>
D: [90] |>
D: [91] |>
D: [92] |>
D: [93] |>
D: [94] |>
D: [95] |>
D: [96] |>
D: [97] |>
D: [98] |>
D: [99] |>
D: [100] |>
D: [101] |>
D: [102] |>
D: [103] |>
D: [104] |>
D: [105] |>
D: [106] |>
D: [107] |>
D: [108] |>
D: [109] |>
D: [110] |>
D: [111] |>
D: [112] |>
D: [113] |>
D: [114] |>
D: [115] |>
D: [116] |>
D: [117] |>
D: [118] |>
D: [119] |>
D: [120] |>
D: [121] |>
D: [122] |>
D: [123] |>
D: [124] |>
D: [125] |>
D: [126] |>
D: [127] |>
D: [128] |>
D: [129] |>
D: [130] |>
D: [131] |>
D: [132] |>
D: [133] |>
D: [134] |>
D: [135] |>
D: [136] |>
D: [137] |>
D: [138] |>
D: [139] |>
D: [140] |>
D: [141] |>
D: [142] |>
D: [143] |>
D: [144] |>
D: [145] |>
D: [146] |>
D: [147] |>
D: [148] |>
D: [149] |>
D: [150] |>
D: [151] |>
D: [152] |>
D: [153] |>
D: [154] |>
D: [155] |>
D: [156] |>
D: [157] |>
D: [158] |>
D: [159] |>
D: [160] |>
D: [161] |>
D: [162] |>
D: [163] |>
D: [164] |>
D: [165] |>
D: [166] |>
D: [167] |>
D: [168] |>
D: [169] |>
D: [170] |>
D: [171] |>
D: [172] |>
D: [173] |>
D: [174] |>
D: [175] |>
D: [176] |>
D: [177] |>
D: [178] |>
D: [179] |>
D: [180] |>
D: [181] |>
D: [182] |>
D: [183] |>
D: [184] |>
D: [185] |>
D: [186] |>
D: [187] |>
D: [188] |>
D: [189] |>
D: [190] |>
D: [191] |>
D: [192] |>
D: [193] |>
D: [194] |>
D: [195] |>
D: [196] |>
D: [197] |>
D: [198] |>
D: [199] |>
D: [200] |>
```

Figure 5. FP-Growth

3.2. Result

From the association patterns obtained, namely 103 patterns, the patterns formed from lending transactions are usually dominated by sustainable or similar books, for example in rule nos. 102 and 103, namely if you borrow morals lecture books, you also borrow morals lectures. Another example is rule no.95, which is if you borrow a book on organic biology 3 and biology, then you borrow an organic biology book 2. In addition to the search for association patterns, a suitability analysis is also carried out between the borrower's study program and the book owner's study program. The results of the analysis are shown in Figure 6. and Figure 7.

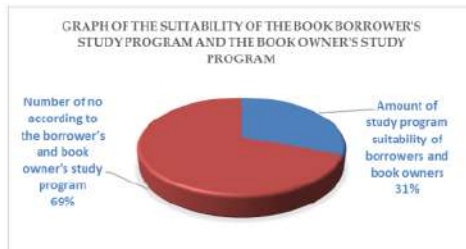


Figure 6. Suitability Results of the Study Program

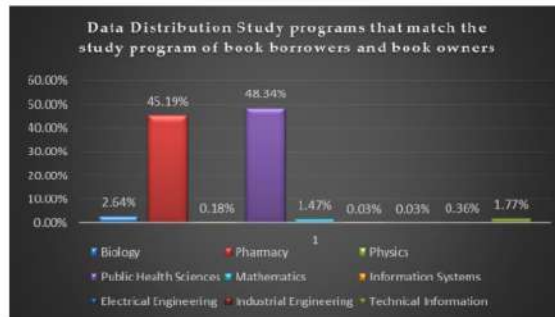
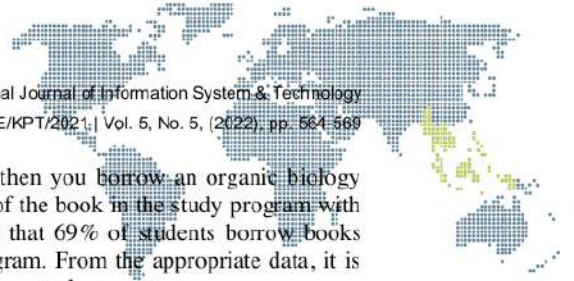


Figure 7. Distribution of Study Programs

4. Conclusion

A program has been made with the python programming language that is able to produce rules which can be used by librarians to determine the distribution of library book lending and the distribution of the suitability of the study program. The resulting pattern is 103 patterns with a support count value of 5 and a confident 10%, this means that the level of book borrowing is still low. From the association pattern obtained, the pattern formed from lending transactions is usually dominated by books that are sustainable or similar, for example in rules nos. 102 and 103, namely if you borrow morals lecture books, you also borrow lectures on morals. Another example is rule no.95, that is, if you





borrow a book on organic biology 3 and biology, then you borrow an organic biology book 2. The results of the analysis of the suitability of the book in the study program with the borrower are 31% that is appropriate, meaning that 69% of students borrow books from the library are not according to the study program. From the appropriate data, it is dominated by the Pharmacy and Public Health Sciences study programs.

## References

- [1] H. Yuliansyah and L. Zahrotun, "Designing web-based data mining applications to analyze the association rules tracer study at university using a FOLD-growth method," *Int. J. Advanced Comput. Res.*, vol. 6, no. 27, pp. 215–221, 2016.
- [2] L. Zahrotun, D. Soyusiawaty, R. S. Pattihua, and U. A. Dahlan, "The Implementation of Data Mining for Association Patterns Determination Using Temporal Association Methods in Medicine Data," pp. 6–11, 2016.
- [3] L. Zahrotun, "Implementation of Data Mining Technique for Customer Relationship Management ( CRM ) on Online Shop Tokodipers . com With Fuzzy C-Means Clustering," in *2017 2nd International Conference on Information TEchnology, Information System, and Electrical Engeneering (ICITISEE) yogyakarta, Indonesia*, 2017, pp. 298–302.
- [4] L. Zahrotun, N. hutami Putri, and A. N. Khusna, "The Implementation of K-Means Clustering Method in Classifying Undergraduate Thesis Titles," in *12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2018.
- [5] L. Zahrotun, "Text Mining for Internship Titles Clustering Using Shared Nearest-Neighbor Method," *Comput. Eng. Appl.*, vol. 6, no. 3, 2017.
- [6] L. Zahrotun, "Comparison Jaccard similarity , Cosine Similarity and Combined Both of the Data Clustering With Shared Nearest Neighbor Method," *Comput. Eng. Appl.*, vol. 5, no. 1, pp. 11–18, 2016.
- [7] M. Kavitha and M. S. T. T. Selvi, "Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons," *Int. journa Comput. Sci. Trends Technol. (IJCST)*, vol. 4, no. 4, pp. 161–164, 2016.
- [8] C. M. Velu, M. Ramakrishnan, V. Somu, and P. Loganathan, P. Vivekanandan, "Efficiency Association Rule for Data Mining," *Int. J. Soft Comput.*, vol. 2, no. 1, pp. 21–36, 2007.
- [9] naufal afif, "Association Rule Mining Peminjaman Buku Perpustakaan Universitas Ahmad Dahlan Menggunakan Algoritma Apriori," 2017.
- [10] G. F. Mandias *et al.*, "Analisa Pola Peminjaman Buku di Pepustakaan Universitas Klabat Menggunakan Algoritma Apriori," *Konf. Nas. Sist. Inf.*, pp. 8–9, 2018.
- [11] J. Han, Kamber.M, Jian Pie, *Data Mining: Concepts and Techniques*. 2011.
- [12] T. Slimani, "Efficient Analysis of Pattern and Association Rule Mining Approaches," *Coll. Comput. Sci. Inf. Technol. , Taif Univ. , KSA*.
- [13] T. Andi and E. Utami, "Association rule algorithm with FP growth for book search," in *3rd Annual Applied Science and Engineering Conference (AASEC 2018)*, 2018, pp. 1–6.
- [14] F. Feng, J. Cho, W. Pedrycz, H. Fujita, and T. Herawan, "Soft Set Based Association Rule Mining," *Knowledge-Based Sist.*, vol. 111, pp. 268–282, 2016.
- [15] C. Angeli, S. K. Howard, M. Jun, J. Yang, and P. A. Kirschner, "Data mining in educational technology classroom research: Can it make a contribution?," *Comput. Educ.*, vol. 113, pp. 226–242, 2017.

# HASIL CEK\_Zahrotun, Jones\_Library, Association Rule, FP-Growth, study program.

## ORIGINALITY REPORT

10%

SIMILARITY INDEX

8%

INTERNET SOURCES

8%

PUBLICATIONS

5%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="https://ojs.stmikpringsewu.ac.id">ojs.stmikpringsewu.ac.id</a> Internet Source	4%
2	Frederick F. Patacsil. "Analysis of Cyberbullying Incidence among Filipina Victims: A Pattern Recognition using Association Rule Extraction", International Journal of Intelligent Systems and Applications, 2019 Publication	1%
3	A A Hidayat, A Rahman, R M Wangi, R J Abidin, R S Fuadi, W Budiawan. "Implementation and comparison analysis of apriori and fp-growth algorithm performance to determine market basket analysis in Breiliant shop", Journal of Physics: Conference Series, 2019 Publication	1%
4	<a href="https://scitepress.org">scitepress.org</a> Internet Source	1%
5	<a href="https://uniassignment.com">uniassignment.com</a> Internet Source	1%



6

[www.iaeme.com](http://www.iaeme.com)

Internet Source

1 %

7

[jurnal-ppi.kominfo.go.id](http://jurnal-ppi.kominfo.go.id)

Internet Source

1 %

8

[doaj.org](http://doaj.org)

Internet Source

1 %

Exclude quotes  On

Exclude matches  < 1%

Exclude bibliography  On

# Fp-Growth Algorithm For Searching Book Borrowing Transaction Patterns And Study Program Suitability

Lisna Zahrotun<sup>1</sup>, Anna Hendri Soleliza Jones<sup>2</sup>

Department of Informatic, Faculty of Industrial Engineering, Universitas Ahmad Dahlan

Ringroad Selatan, Kampus 4 UAD, Yogyakarta 55191

e-mail: lisna.zahrotun@tif.uad.ac.id<sup>1\*</sup>, annahendri@tif.uad.ac.id<sup>2</sup>

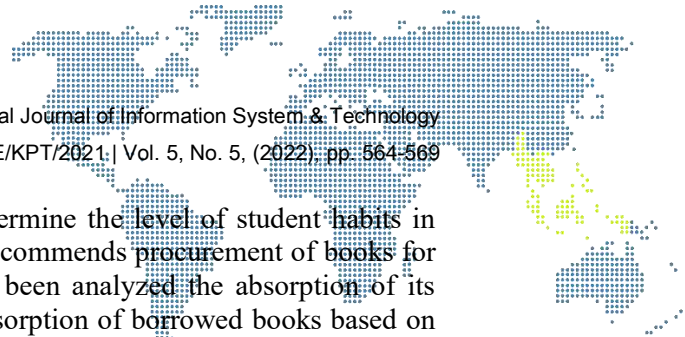
## Abstrak

The current development of data has reached a sizeable amount. This is due to the development of the world of information technology which consists of data in it. One technique that can handle abundant data is data mining. Data mining methods are widely used to perform large amounts of data analysis. In the academic field, analysis can be used to determine the patterns of students and lecturers. Whereas in library transactions, analysis can be carried out to determine the patterns of existing book borrowing. This is done to determine the tendency of students with certain study programs to borrow any uku transactions. In this study, the aim of this research is to analyze the patterns of borrowing books from the Ahmad Dahlan University library, which includes borrowing transaction data and the book owner's study program. In addition, in this study, a percentage analysis of the suitability of the book borrower study program and the book owner's study program was also carried out. The stages in this research include data collection, data cleaning, data selection, data transformation, searching for association patterns using the FP-Growth method and pattern evaluation. The test used in this research is the lift ratio. The results of this study are publications in international journals that are in the draft process. Apart from that, the results of this study provide information on the analysis of patterns of lending books in libraries using the FP-Growth method. The resulting pattern is 103 patterns with a support count value of 5 and a confident 10% with the 2 itemset rule, this means that the level of book borrowing is still low. While the results of the analysis of the suitability of books in the study program with the borrower were 31% in accordance with the study program, namely Pharmacy and Public Health Sciences, meaning that there were 69% of students who borrowed books from the library that were not in accordance with their study program.

**Keywords:** Library, Association Rule, FP-Growth, study program.

## 1. Introduction

The library is a collection of information that can be accessed by the public to find references or borrow materials / information. The information provided by the library can be physical or digital. One of the functions of the library is a place for borrowing books which is a provider of information and knowledge. The quality of the library can be seen from how well the book lending and borrowing transaction function is carried out. Based on library interviews, a university in Yogyakarta also has more than 24,397 collections of book titles totaling 67,453 copies. Transactions that occur in the library every day reach 100 transactions, all of which have been recorded in the system. However, the large number of books means that there are also many books that are not actively borrowed by students. Therefore, it takes a step to find book borrowing patterns to determine the level of student habits in borrowing books. Apart from that every days reached 100 transactions which have all been recorded in the system. However, the large number of books means that there are also many books that are not actively borrowed by students. Therefore, it



takes a step to find book borrowing patterns to determine the level of student habits in borrowing books. In addition, each study program recommends procurement of books for each year. However, this available book has never been analyzed the absorption of its borrowing. Therefore it is important to know the absorption of borrowed books based on the borrower's study program and the book owner's study program. In previous research, we have conducted research related to data mining, namely searching for patterns of alumni data associations [1], drug data patterns [2], grouping customer data [3]. Apart from data mining, research on text mining is also carried out [4][5][6].

One way to solve the problems in the library is by implementing the Association rule. association rule is a data mining technique used to find associative rules between a combination of items. Frequent Pattern Growth (FP-Growth) is an a priori algorithm that can be used to determine the most frequent itemsets in a set of data. FP-Growth is faster than the Apriori algorithm [7]. By using the FP-Tree concept, the FP Growth algorithm also has a good ability to look for association patterns [8]. Several studies have been conducted to find out patterns of lending library books. As research conducted by [9] and on other studies conducted [10]. Both studies have been carried out and both are still using the a priori algorithm, which is basically the process of calculating the frequent itemset is still relatively slow, so this study has determined the eclat algorithm so that the frequent itemset calculation process is faster. In addition, research conducted at the library has not implemented the suitability of study program books with borrower study programs and still uses book IDs for pattern search calculations.

## 2. Research Methodology

### 2.1. Data Mining

Data mining is a process of finding important knowledge from large collections of data that have been stored in databases [11]. Such knowledge cannot be found in simple ways. Data mining uses certain techniques (methods) to find knowledge including classification, clustering and linkage analysis. According to [12] Data mining tasks can be grouped into two categories namely:

- a) Predictive mining is concerned with classification methods, regression and deviation detection.
- b) Descriptive mining will derive and investigate patterns (correlation, trend, cluster) of important relationships from data sets. The main method of descriptive mining is clustering, association rules and sequential mining.

### 2.2. FP-Growth

FP-Growth is an improvement on the a priori algorithm. This algorithm is used to define the most frequent itemset in a data set [13]. The FP-Growth algorithm uses the concept of a development tree in searching for frequent itemsets. The characteristic of the FP-Growth algorithm is that the data structure used is a tree called the FP-Tree. By using the FP-Tree, the FP-Growth algorithm can directly extract itemset from the FP-Tree. Excavation of frequent itemset using the FP-Growth algorithm will be carried out by generating a data structure tree or called the FP-Tree [14][15].

### 2.3. Factor of Interest

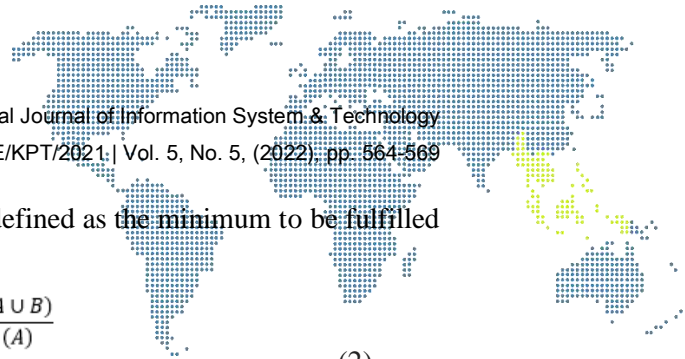
Support is an occurrence rule for all transactions in the dataset [11]. Minimum support is a value that must be met by the rules:

$$\text{Support}(A \rightarrow B) = P(A \cup B) = \frac{\sum A \cup B}{\sum \text{Transaction}} \quad (1)$$

### 2.4. Confidence

Confidence is the opportunity for an item to appear together with other items to appear. If it is formed with notation in  $A \rightarrow B$ , then trust is defined as how often B occurs





when A also occurs. Minimum Confidence value is defined as the minimum to be fulfilled by the rules.

$$Confidence(A \rightarrow B) = P(A|B) = \frac{P(A \cup B)}{P(A)} = \frac{support\ count(A \cup B)}{support\ count(A)} \quad (2)$$

### 2.5. Lift Rasio

Lift Ratio is used to find out the correlation between the items in the rule. If the value of the lift rule is  $> 1$  then Positively correlated, if the value of the lift rule  $< 1$  then negatively correlated, if the value of the lift rule  $= 1$  then it's independent (not combined).

$$Lift(A \rightarrow B) = P(A|B) = \frac{P(A \cup B)}{P(A).P(B)} = \frac{confidence(A \cup B)}{support(A)} \quad (3)$$

### 2.6. Frekuent Pattern Growth

FP-Growth is one of the algorithms used to solve the Association Rule case. This algorithm has two steps. The first two steps, compression is performed on the database based on frequently occurring items to create a Frequent Pattern Tree (FP-Tree). Second, separation is performed on the database results into a compressed database condition. FP-Tree is a special feature that distinguishes the FP-Growth algorithm from the a priori algorithm. FP-Tree has two features, first starting from the root which is named "null". Then from the root to form a sub-tree consisting of certain items. Second, each node contains three important information, namely the item label (indicates the type of item (item ID) it is represented by the node), support (the number shows the number of transaction paths through that node or also known as frequency), connecting pointers (link nodes) as a liaison. Between vertices and items in each path between cells, marked with a line connecting the pointer arrows.

## 3. Result and Discussion

The results of this study are the association pattern and the percentage of suitability of book borrowers with the book owner's study program at a university library. Before the pattern search is carried out, the first step taken is taking the data of the library book lending transaction that will be processed. The data that has been successfully displayed is then carried out by a process of selecting and transforming the data. The results of the selection and data transformation then carried out a pattern search using the FP Growth method. After finding several patterns, the next step the borrower's suitability pattern and graph with the book owner's program of study. An image of the research is to check the correlation of the patterns produced using the lift ratio. And the final step is to display the results of steps is shown in Figure 1.

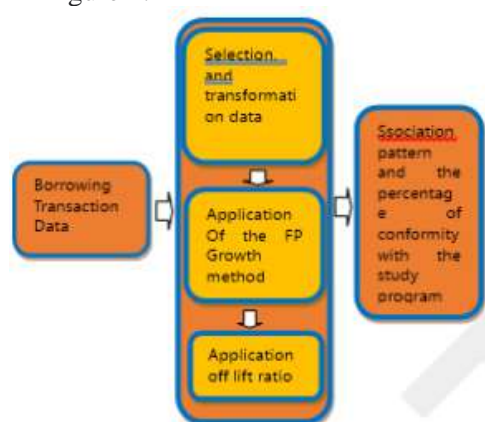
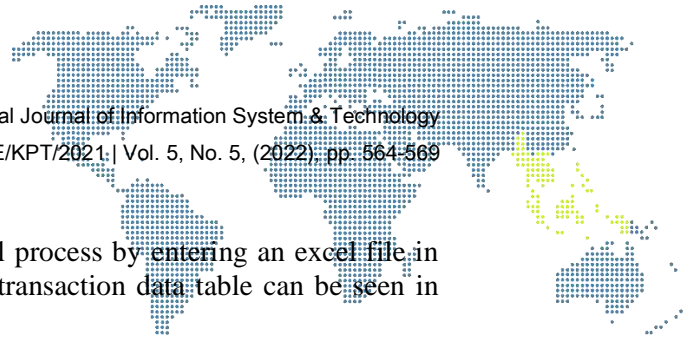


Figure 1. Research Diagram



### 3.1. Implementation

The first step is data retrieval which is the initial process by entering an excel file in the form of initial book loan transaction data. The transaction data table can be seen in Figure 2.

**Uploaded Data**

	0	1	2	3
6	6	0748V2018IKM3	074/B/V/2018/IKM/3/C.7	Buku Ajar Kesehatan Ke...
7	7	103BI2017LPSI	103/B/II/2017/LPSI/3/C...	Tuntunan Ibadah Prakti...
8	8	12198V2010FAM	1219/B/V/2010/FAR/3/C...	Taksonomi Umum Dasar-B...
9	9	1308V2007IKM	130/B/V/2007/IKM/3	Buku Saku Hematologi
10	10	0420I12017FAR	042/B/II/2017/FAR/3/C.9	Buku Ajar Vogel Kimia ...
11	11	052HVII1966FAR	052/H/VII/1966/FAR/3	Biologi Jilid 2
12	12	2948IV2000IKM	294/B/IV/2000/IKM/3	Perilaku Organisasi Ji...
13	13	7088VII2005IKM	708/B/VII/2005/IKM/3	Manajemen Personalia d...
14	14	2218XI2002TK	221/B/XI/2002/TK/3	Introduction to Chemi...
15	15	0908II2017LPSI	090/B/II/2017/LPSI/3/C...	Kuliah Umum Hadis
16	16	0378IV2010NP3	037/B/IV/2010/NP/3/C.2	CITIZEN 4.0 : Menajak...

**Figure 2. Load Data**

The next stage, namely data selection is carried out before the information mining stage. The selected data is stored in the file. at this stage remove unused variables. The data that has been successfully selected can be seen in Figure 3.

In [15]: df

Out[15]:

	0	1	2	3	4	5	6	7	8	9	10
6	1	10030X2017FAR	1003/B/X/2017/FAR/3/C.14	Kimia Farmasi Analisa	1900020206	Aula Hazma Putri	31-12-2018	-	None	Kembali dan Tertarik	Formasi
7	2	0548VI12017IKM	0548/B/VI/2017/IKM/3/C.4	Metode Penelitian Kuantitatif dan R&D	1708040040	SITI BUSTANI FAUDAH	31-12-2018	-	None	Kembali dan Tertarik	Manajemen Pendidikan SD
8	3	3808II2004IKM	3808/B/II/2004/IKM/3	Etika	1800029003	NORMLINA FEBRIANA	30-12-2018	-	None	Kembali dan Tidak Tertarik	Ilmu Kesehatan Masyarakat
9	4	3808II2004IKM	3808/B/II/2004/IKM/3	Etika	1800029003	NORMLINA FEBRIANA	30-12-2018	-	None	Kembali dan Tidak Tertarik	Ilmu Kesehatan Masyarakat
10	5	1308V2010IKM	1308/B/V/2010/IKM/3/C.5	Kesehatan Lingkungan dan K2	1608044207	Liza Feoriana Fumana	29-12-2018	-	None	Kembali dan Tidak Tertarik	Pasca Sarjana Psikologi Sains

**Figure 3. Selected Data**

The next stage, namely this transformation is carried out by changing the data variable into valid or ready for the data mining process. The results can be seen in Figure 4

**Data Transformation**

	Inventaris	Prodi
0	Formasi	Formasi
1	Ilmu Kesehatan Masyarakat	Manajemen Pendidikan S2
2	Ilmu Kesehatan Masyarakat	Ilmu Kesehatan Masyarakat
3	Ilmu Kesehatan Masyarakat	Ilmu Kesehatan Masyarakat
4	Ilmu Kesehatan Masyarakat	Pasca Sarjana Psikologi Sains
5	Ilmu Kesehatan Masyarakat	Pasca Sarjana Psikologi Sains
6	Lembaga Pengembangan Studi Islam	Biologi
7	Farmasi	Biologi
8	Ilmu Kesehatan Masyarakat	Biologi
9	Farmasi	Teknik Kimia - S1
10	Farmasi	Pendid. Biologi

**Figure 4. Data Transformation**

The final step is Result FP-growth to display the results of the association rules where there is an input form to input minimum confidence and minimum support with the minimum confidence 0.7 and minimum support conditions 8. The results of the FP-growth result can be seen in Figure 5.

```

In [14]: import pyfigrowth

In [15]: patterns = pyfigrowth.find_frequent_patterns(iris_book.values(), 5)
patterns

Out[15]: [{"book1": "Banyak Jantung Koroner dan Serangan Jantung Pencegahan Penyembuhan Rehabilitasi Penderita Bagi Masyarakat Umum", "support": 5},
{"book1": "Manajemen Personalia dan Sumberdaya Manusia", "support": 5},
{"book1": "Masa Depan Anak Yang Terate Di Indonesia Dan Di Dunia", "support": 5},
{"book1": "Kesamaran Ibu dan Anak (KIA) Dalam Millenium Development Goals (MDG)", "support": 5},
{"book1": "Kesehatan Tradisional", "support": 5},
{"book1": "Tali Duri dan Penyakit Kerebohan Masyarakat", "support": 5},
{"book1": "Drug Information Handbook A Clinically Relevant Resource for All Healthcare Professionals", "support": 5},
{"book1": "Prinsip-prinsip Klinis Medisil Jilid 1", "support": 5},
{"book1": "Mikrobiologi Dasar", "support": 5},
{"book1": "Pharmacopoeia U", "support": 5},
{"book1": "Tanya Jawab Igama 2", "support": 5},
{"book1": "Malaria epidemiologi, Patogenesis, manifestasi klinis, & penanganannya", "support": 5},
{"book1": "Lingkungan Hidup dan Pencemaran Bahayanya dengan Toksikologi Senyawa Organik", "support": 5},
{"book1": "Tumbuh Kembang Anak", "support": 5},
{"book1": "Anestesi dan Anestesi Kehamilan", "support": 5},
{"book1": "Toksikologi Umum Pengantar", "support": 5},
{"book1": "Materia Medica Indonesia Jilid II", "support": 5},
{"book1": "Tinjauan Klinis Hasil Pemeriksaan Laboratorium", "support": 5},
{"book1": "Mikrobiologi", "support": 5}
    
```

Figure 5. FP-Growth

**3.2. Result**

From the association patterns obtained, namely 103 patterns, the patterns formed from lending transactions are usually dominated by sustainable or similar books, for example in rule nos. 102 and 103, namely if you borrow morals lecture books, you also borrow morals lectures. Another example is rule no.95, which is if you borrow a book on organic biology 3 and biology, then you borrow an organic biology book 2. In addition to the search for association patterns, a suitability analysis is also carried out between the borrower's study program and the book owner's study program. The results of the analysis are shown in Figure 6. and Figure 7.

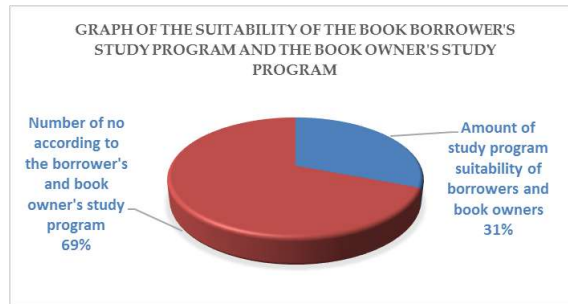


Figure 6. Suitability Results of the Study Program

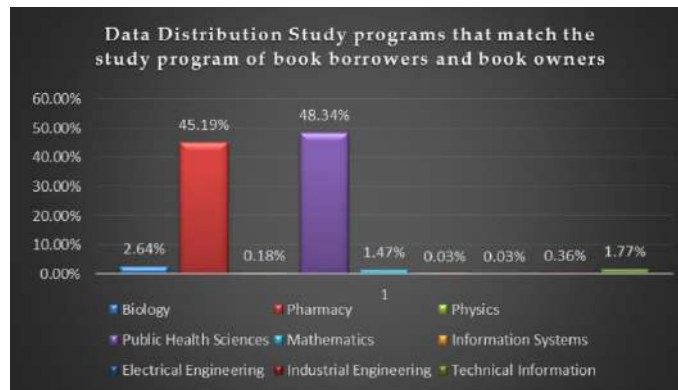
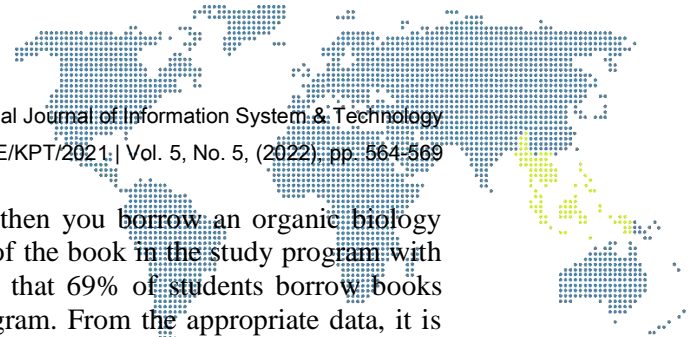


Figure 7. Distribution of Study Programs

**4. Conclusion**

A program has been made with the python programming language that is able to produce rules which can be used by librarians to determine the distribution of library book lending and the distribution of the suitability of the study program. The resulting pattern is 103 patterns with a support count value of 5 and a confident 10%, this means that the level of book borrowing is still low. From the association pattern obtained, the pattern formed from lending transactions is usually dominated by books that are sustainable or similar, for example in rules nos. 102 and 103, namely if you borrow morals lecture books, you also borrow lectures on morals. Another example is rule no.95, that is, if you





borrow a book on organic biology 3 and biology, then you borrow an organic biology book 2. The results of the analysis of the suitability of the book in the study program with the borrower are 31% that is appropriate, meaning that 69% of students borrow books from the library are not according to the study program. From the appropriate data, it is dominated by the Pharmacy and Public Health Sciences study programs.

## References

- [1] H. Yuliansyah and L. Zahrotun, "Designing web-based data mining applications to analyze the association rules tracer study at university using a FOLD-growth method," *Int. J. Advanced Comput. Res.*, vol. 6, no. 27, pp. 215–221, 2016.
- [2] L. Zahrotun, D. Soyusiawaty, R. S. Pattihua, and U. A. Dahlan, "The Implementation of Data Mining for Association Patterns Determination Using Temporal Association Methods in Medicine Data," pp. 6–11, 2016.
- [3] L. Zahrotun, "Implementation of Data Mining Technique for Customer Relationship Management ( CRM ) on Online Shop Tokodipers . com With Fuzzy C-Means Clustering," in *2017 2nd International Conferenceseon Information TEchnology, Information System, and Electrical Engeenering (ICITISEE) yogyakarta, Indonesia, 2017*, pp. 298–302.
- [4] L. Zahrotun, N. hutami Putri, and A. N. Khusna, "The Implementation of K-Means Clustering Method in Classifying Undergraduate Thesisi Titles," in *12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2018.
- [5] L. Zahrotun, "Text Mining for Internship Titles Clustering Using Shared Nearest-Neighbor Method," *Comput. Eng. Appl.*, vol. 6, no. 3, 2017.
- [6] L. Zahrotun, "Comparison Jaccard similarity , Cosine Similarity and Combined Both of the Data Clustering With Shared Nearest Neighbor Method," *Comput. Eng. Appl.*, vol. 5, no. 1, pp. 11–18, 2016.
- [7] M. Kavitha and M. S. T. T. Selvi, "Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons," *Int. journa Comput. Sci. Trends Technol. (IJCST)*, vol. 4, no. 4, pp. 161–164, 2016.
- [8] C. M. Velu, M. Ramakrishnan, V. Somu, and P. Loganathan, P. Vivekanandan, "Efficiency Association Rule for Data Mining," *Int. J. Soft Comput.*, vol. 2, no. 1, pp. 21–36, 2007.
- [9] naufal afif, "Association Rule Mining Peminjaman Buku Perpustakaan Universitas Ahmad Dahlan Menggunakan Algoritma Apriori," 2017.
- [10] G. F. Mandias *et al.*, "Analisa Pola Peminjaman Buku di Pepustakaan Universitas Klabat Menggunakan Algoritma Apriori," *Konf. Nas. Sist. Inf.*, pp. 8–9, 2018.
- [11] J. Han, Kamber. M., Jian Pie, *Data Mining: Concepts and Techniques*. 2011.
- [12] T. Slimani, "Efficient Analysis of Pattern and Association Rule Mining Approaches," *Coll. Comput. Sci. Inf. Technol. , Taif Univ. , KSA*.
- [13] T. Andi and E. Utami, "Association rule algorithm with FP growth for book search," in *3rd Annual Applied Science and Engineering Conference (AASEC 2018)*, 2018, pp. 1–6.
- [14] F. Feng, J. Cho, W. Pedrycz, H. Fujita, and T. Herawan, "Soft Set Based Association Rule Mining," *Knowledge-Based Sist.*, vol. 111, pp. 268–282, 2016.
- [15] C. Angeli, S. K. Howard, M. Jun, J. Yang, and P. A. Kirschner, "Data mining in educational technology classroom research : Can it make a contribution?," *Comput. Educ.*, vol. 113, pp. 226–242, 2017.