

PAPER • OPEN ACCESS

Association rule algorithm with FP growth for book search

To cite this article: T Andi and E Utami 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **434** 012035

View the [article online](#) for updates and enhancements.

You may also like

- [Personalized Recommendation Algorithm for books and its implementation](#)
Li Chun-mei, Ma yi-han, Pi Wei et al.
- [Implementation of ECLAT Algorithm Technology: Determining Books Borrowing Pattern in University library](#)
Lisnawita and Mariza Devega
- [Efficient quantum algorithm for all quantum wavelet transforms](#)
Mohsen Bagherimehrab and Alán Aspuru-Guzik



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Association rule algorithm with FP growth for book search

T Andi* and E Utami

Magister of Informatics Engineering Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

*anditri14@gmail.com

Abstract. Collection of books in the library is sometimes confusing visitors who want to borrow library books. Data mining techniques have been widely used to resolve existing problems by applying the algorithm of association rule Algorithm using FP Growth to find the rules of the Association that is formed from the dataset loaning books. By utilizing the visit transaction data, the library will be seeking information about what books are often borrowed by the student and the linkages between each of the parties so that the lending library itself can do the preparation books in accordance with the level of support and confidence values. One of the rules of the Association that is formed is if someone borrows the book Psikologi Perkembangan then he will borrow the book Psikometri with a value of 0.2 and support 100% confidence rating which is the highest confidence value rules. Then create an application that can show the location of the book more specifically so it can facilitate the students in searching.

1. Introduction

Data is always a part of every activity we do today, data growth is inevitable. The size starts from byte to petabyte. To grow millions, even billions of data stored in electronic data storage, the right tools are needed. Data processing here aims to produce information that can assist in making management decisions. From this base, the concept of data mining was born. Data mining is the process by which intelligence methods are used to generate data patterns. XYZ has a library of visitor statistics of 50 people per day in 2018. The number of books available reaches 2,000 books. For visitors looking for references that are addressed to the topic, the book search system already available in this library is the main target. Because in search of references, it is usually required more than one book, then the book recommendations of the system facilities will greatly help visitors.

2. Data warehouse

The database is a set of interconnected data groups that are arranged in such a way that can later be reused quickly and easily. The database is part of the system data. A database is a collection of multiple files that have a link between one file with another file to form data. While the database system is the foundation or basic framework consisting of system and the system used to serve certain applications in order to achieve the goal [1] [2] [3].



3. Method

3.1. Data mining

The method of excavating Association rules (association mining rules) is one method in the concept of data mining, which aims to find interesting patterns in a number of transactions. Data Mining is a term used to describe the discovery of knowledge in database [4] [5]. Data Mining is the process of using statistical, mathematical, artificial intelligence, and machine learning techniques to identify and extract useful information and knowledge from various related database [6] [7].

3.2. FP Growth algorithm

FP Growth algorithm is an improvement from the a priori algorithms in terms of speed of execution so that the shortcomings of the a priori algorithms improved by FP-Growth algorithm. Frequent Pattern Growth (FP-Growth) is one of the alternative algorithm that can be used to define the set of data that appears most often (frequent itemset) in a set of data [8]. On the a priori algorithm required generate candidate to get frequent itemsets. However, at the FP-Growth algorithm generate candidate was not done because the FP-Growth using the concept of the development tree in search of frequent itemsets. Characteristics of FP-Growth algorithm is the data structure used is a tree called FP-Tree. By using FP-Tree, FP-Growth algorithm can directly extract frequent itemset of FP-Tree. The excavation of the frequent itemset by using FP-Growth algorithm will be done by way of evoking data structure tree or called FP-Tree [9] [10].

3.3. Factor of interest

Support is a chance occurrence rule to the overall transaction in the dataset [11]. The minimum support value is defined as the minimal support that must be met by the rule:

$$Support(A \rightarrow B) = P(A \cup B) = \frac{\sum A \cup B}{\sum Transaksi}$$

3.4. Confidence

Confidence is a chance occurrence an item appear along with other items appear. If it forms with the notation in $A \rightarrow B$, then confidence defined as how often B appears when A also appears. Minimum confidence value is interpreted as a minimum to be met by rule [12] [13].

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

3.5. Lift ratio

Lift ratio is used to find out the correlation between the items in the rule. If the value of lift rule > 1 then it has positive correlation, if the value of lift rule < 1 then it has negative correlation, if the value of lift rule $= 1$ then it is independent (not coupled) [14].

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

3.6. Frequent pattern growth (FP-Growth)

FP-Growth is one of the algorithms used to solve the case of the Association Rule. This algorithm has two stages: first, the compression is performed against the database on the basis of the items that often appear to make Frequent Pattern Tree (FP-Tree). Second, the separation is performed against the database results into the form of compressed database conditionals [15] [16].

3.7. FP-Tree

FP-Tree is a special trait that distinguishes between the FP-Growth algorithm with a similar algorithm, a priori [17]. FP-Tree has two characteristics, first, starting from a root that is named "null". Then from the root to form sub-tree consisting of specific items and a frequent table header. Second, each node contains three important information i.e. the label item (indicates the item type (item ID) that is represented by the node), support (count shows the number of transaction path through that node or also known as as frequency), connecting the pointer (node link) as a liaison among vertices with items at once inter-cell trajectory, marked with a line connecting the pointer arrow disconnect – disconnect [18] [19] [20].

4. Results and discussion

This research will find the value of support and trust of the book loan in the library XYZ. These book loan transaction data will be mined by applying association rules so as to bring new association rules that serve to provide useful information to libraries such as sysops that know what books are usually borrowed by members, so it can be done with better preparatory books to facilitate the search book.

Job steps of the algorithm Association Rule is as follows:

- Determine the minimum value of the amount of support and confidence, which in this study the minimum amount of support support = 4, while the minimum quantity = 50% confident.
- Scan the transaction base lending library books to get the 1-item set, then grab the frequency candidate to meet the minimum support amount. While it does not meet the minimum support removed or deleted.
- Check whether $k + 1$ item set is from the existing item set, then discard or clean the candidate data of the subset does not meet the support amount.
- Calculate the frequency, take any frequency meet and finish if there is no eligible policy.
- Association Rules are generated by selection that meets minimum support amount and Analytical trust

4.1. Association rule

Before performing data mining with Association Rule method, please be prepared to be on mining data where data used is loaning books transaction data by using fictitious data. It is intended to give you an idea how to do data mining Association rules resulting in (Association Rules). To do the excavation of loaning books transaction data, then the steps that must be performed are:

- The first step is in controlling the magnitude of quantity minimum support and confidence. Where in this case the set quantity = quantity of 3 and support confidence = 50%.
- The second step is to arrange all the frequent item set i.e. item set which has the minimum support = 4 set earlier.

For analysing, it starts by discussing any frequent 1-item set, then each selected item has the minimum support = 4. Items that meet the minimum support are called frequent 1-item set. List of frequent 1-item set can be seen in the table below:

Table 1. List 1-frequent item set.

No	Item set	Support
1	<i>Pengantar Psikologi</i>	4
2	<i>Psikologi Perkembangan</i>	9
3	<i>Psikologi abnormal</i>	7
4	<i>Psikometri</i>	5
5	<i>Psikologi Sosial</i>	3
6	<i>Psikologi Klinis</i>	4
7	<i>Biopsikologi</i>	7
8	<i>Psikologi Pendidikan</i>	3

From the table above it can be seen that the eligible minimum support or eligible to be 2-Item is as shown in table 3 below:

Table 2. The list of Prospective for the Frequent 2-Item set.

No	Itemset	Support
1	<i>Pengantar Psikologi, Psikologi Perkembangan</i>	1
2	<i>Pengantar Psikologi, Psikologi Abnormal</i>	1
3	<i>Pengantar Psikologi, Psikometri</i>	0
4	<i>Pengantar Psikologi, Psikologi Sosial</i>	1
5	<i>Pengantar Psikologi, Psikologi Klinis</i>	0
6	<i>Pengantar Psikologi, Biopsikologi</i>	0
7	<i>Pengantar Psikologi, Psikologi Pendidikan</i>	1
8	<i>Psikologi Perkembangan, Psikologi Abnormal</i>	1
9	<i>Psikologi Perkembangan, Psikometri</i>	4
10	<i>Psikologi Perkembangan, Psikologi Sosial</i>	0
11	<i>Psikologi Perkembangan, Psikologi Klinis</i>	0
12	<i>Psikologi Perkembangan, Biopsikologi</i>	6
13	<i>Psikologi Perkembangan, Biopsikologi</i>	1
14	<i>Psikologi abnormal, Psikometri</i>	0
15	<i>Psikologi abnormal, Psikologi Sosial</i>	1
16	<i>Psikologi abnormal, Psikologi Klinis</i>	5
17	<i>Psikologi abnormal, Biopsikologi</i>	0
18	<i>Psikologi abnormal, Psikologi Pendidikan</i>	0
19	<i>Psikometri, Psikologi Sosial</i>	1
20	<i>Psikometri, Psikologi Klinis</i>	2
21	<i>Psikometri, Biopsikologi</i>	5
22	<i>Psikometri, Psikologi Pendidikan</i>	1
23	<i>Psikologi Sosial, Psikologi Klinis</i>	1
24	<i>Psikologi Sosial, Biopsikologi</i>	1
25	<i>Psikologi Sosial, Psikologi Pendidikan</i>	0
26	<i>Psikologi Klinis, Biopsikologi</i>	1
27	<i>Psikologi Klinis, Psikologi Pendidikan</i>	0
28	<i>Biopsikologi, Psikologi Pendidikan</i>	

From the table above it can be seen that a minimum qualified support or are eligible to become frequent 2-Item set is as shown in table 3 below:

Table 3. List 2-frequent item set.

No	Itemset	Support
1	<i>Psikologi Perkembangan, Psikometri</i>	4
2	<i>Psikologi Perkembangan, Biopsikologi</i>	6
3	<i>Psikologi abnormal, Psikologi Klinis</i>	5
4	<i>Psikometri, Psikologi Klinis</i>	5

Next the Item set list is often above 1-made into a list of candidates for frequent 3-Itemsets as shown in the following table 4:

Table 4. List 3-frequent itemset.

No	Itemset	Support
1	<i>Psikologi Perkembangan, Psikometri, Biopsikologi</i>	4
2	<i>Psikologi Perkembangan, Psikometri, Psikologi Klinis</i>	0
3	<i>Psikologi Perkembangan, Psikometri, Psikologi Abnormal</i>	0
4	<i>Psikologi Perkembangan, Biopsikologi, Psikologi Klinis</i>	0
5	<i>Psikologi Perkembangan, Biopsikologi, Psikologi Abnormal</i>	0

From the results in the table above it can be seen that the right to be a frequent 3-Itemset is as shown in table 5 below:

Table 5. List 3-frequent itemset

No	Itemset	Support
1	<i>Psikologi Perkembangan, Psikometri, Biopsikologi</i>	4

From the table above it can be seen that the right to be a frequent 3-Itemset, if someone borrows Developmental Psychology then he borrows Psikometri, it will be obtained the value of confidence = $4/9 = 0.44 = 44\%$. From the list of candidates of the above Association rules, then the rules of the Association will be selected who meet the minimum support and minimum confidence to become the rules of the Association. Where the minimum support and minimum confidence = $4 = 50\%$ as it has been set previously.

5. Conclusion

Association rule mining is used to see the interesting relationship between a number of lending books. In the process of association rule mining there are two search patterns and calculate the force of a rule. From the results of the data mining library book lending transaction with a minimum transaction limit of 4 minimum transaction and confidence of 50% form 21 rules of the Association. One of the rules of the Association that is formed is if someone borrows the book Psikologi Perkembangan then he will borrow the book Psikometri with a value of 0.2 and support 100% confidence rating which is the highest confidence value rules.

References

- [1] J Han, J Pei and M Kamber 2011 Data Mining: Concepts and Techniques *Elsevier Science*
- [2] L W Santoso and Yulia 2017 Data Warehouse with Big Data Technology for Higher Education *Procedia Comput. Sci.* **124** pp. 93–99
- [3] D T Larose 2014 *Discovering Knowledge in Data: An Introduction to Data Mining* (Wiley)
- [4] E Turban, J E Aronson and T P Liang 2005 *Decision Support Systems and Intelligent Systems* (Prentice Hall)
- [5] M R Abdellah, H A Mohamed, K S Badran and M B Senousy 2017 Integrated Association Rules Complete Hiding Algorithms *Adv. Electr. Electron. Eng.* **15**(2) pp. 192–202
- [6] G Prati, M De Angelis, V M Puchades, F Fraboni and L Pietrantoni 2017 Characteristics of cyclist crashes in Italy using latent class analysis and association rule mining *PLoS One* **12**(2) pp. 1–28
- [7] A S Aribowo and N H Cahyana 2015 Feasibility study for banking loan using association rule mining classifier *Int. J. Adv. Intell. Informatics* **1**(1) pp. 41–47
- [8] V Patil, R Vasappanavara and T Ghorpade 2016 Securing association rule mining with FP growth algorithm in horizontally partitioned database *Icccm*
- [9] F Feng, J Cho, W Pedrycz, H Fujita and T Herawan 2016 Soft set based association rule mining *Knowledge-Based Syst.* **111** pp. 268–282
- [10] C Angeli, S K Howard, J Ma, J Yang and P A Kirschner 2017 Data mining in educational technology classroom research: Can it make a contribution? *Comput. Educ.* **113** pp. 226–242
- [11] J M Adamo 2012 Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms (New York: Springer)
- [12] K Khurana and S Sharma 2013 A comparative analysis of association rule mining algorithms *Int. J. Sci. Res. Publ.* **3**(5) p. 0
- [13] M Kaur and S Kang 2016 Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining *Procedia Comput. Sci.* **85** pp. 78–85
- [14] B Noh, J Son, H Park and S Chang 2017 In-Depth Analysis of Energy Efficiency Related Factors in Commercial Buildings Using Data Cube and Association Rule Mining *Sustainability* **9**(11) p. 2119
- [15] M Salles 2015 Decision-Making and the Information System (Wiley)
- [16] A C McLucas 2003 Decision Making: Risk Management, Systems Thinking and Situation

Awareness *Argos Press*

- [17] V L Sauter 2014 Decision Support Systems for Business Intelligence (Wiley)
- [18] R Setiono, W K Leow and J M Zurada 2002 Extraction of rules from artificial neural networks for nonlinear regression *IEEE Trans. Neural Netw.* **13**(3) pp. 564–577
- [19] K Rajendra Prasad 2017 Optimized high-utility itemsets mining for effective association mining paper *Int. J. Electr. Comput. Eng.* **7**(5) pp. 2911–2918
- [20] S O Fageeri, R Ahmad and H Alhussian 2016 A performance analysis of association rule mining algorithms in 2016 *3rd International Conference on Computer and Information Sciences (ICCOINS)* pp. 328–333