Association pattern of students thesis examination using fp-growth algorithms

Ika Arfiani [a,1,*], Herman Yuliansyah [a,2], Tia Purwantias [a,3]

[a] Department of Informatics, Faculty of Industrial Technology, Universitas Ahmad Dahlan, Indonesia
[1] *ika.arfiani@tif.uad.ac.id* *; [2] *herman.yuliansyah@tif.uad.ac.id; [3] tiastias0708@gmail.com*
* *corresponding author*

ABSTRACT

The thesis examination is the final project for students to graduate from their majors. This thesis researches scientific work between a student and a supervisor in finding solutions to a problem. In the thesis examination, students must present their research results to be criticized by the examiner. This article aims to analyze the association pattern of student thesis examinations at a private university. Although the thesis's implementation has been carried out following procedures, to determine the composition of the board of examiners needs to be analyzed by examining the pattern of relationships between research topics, supervisors, and examiners. This study uses 448 data and uses FP-Growth Algorithms to find the rules. The research methodology starts from preparing the Dataset, cleansing data, selecting data, loading data into applications, transforming data, itemset frequencies, forming patterns, and analyzing rules. This study found 145 patterns of association rules with a minimum support value = 4 and a minimum trust value = 50%. The association rule pattern of 77.78% is under scientific group data. The benefits of the association pattern produced in this study can determine the composition of the examiners on the student thesis examination according to the research topic and scientific field of the examiners.

Keywords:
Data Mining,
 Association Rules Mining,
FP-Growth Algorithms,
Students Thesis
Examination,
Data Patterns

I. Introduction

Association rule mining is a technique in data mining instead of classification, clustering, and prediction. By using data mining, the data scientist can find new insight from the data. Data mining can implement in several areas, i.e., discovering client behavior patterns that can use to address the marketing strategies[1], discovering undergraduate student data[2], and extracting facts from medical data[3] by association rule mining. Data mining also uses to identify toddlers nutritional status[4], extraction risk factor from the highly ranked association rules for early childhood caries[5], analysis of the relationship among patent documents[6], analysis of ozone condition[7], [8], and determining the pharmacies drug to prevent a drug mistake[9].

Educational data mining is an interesting topic to discuss. The main goal of implementing data mining in the educational area is to use experience and new insight to improve the quality of education[10] and manage new courses properly[11]. Data mining also can use to prevent educational risk and educational opportunities i.e., student drop-out[12]–[16], duration of study[17], [18], learning behaviors[19]–[21], students outcome[22], [23] and student performance[10], [24]–[26].

In the previous research, the FOLD-growth method was used to analyze Alumni data's frequent pattern relation between time to get a job, study duration, age, English skill, field skill, grade point average, and the first salary [27]. This article will discuss the unique implementation of data mining for the educational field. By using association rules mining, we interest in analyzing student thesis examination to find meaningful association patterns. The thesis is a written scientific work prepared by students based on the results of the research problem. This research was carried out carefully with the guidance of a supervisor and research work from students. The results of this study are expected to contribute to knowledge and society. Every university and department has its procedure to provide student thesis examination. In general, every student thesis examination consists of four attributes, i.e., student, supervisor, examiner, and topic research. There is no problem in implementing student thesis examination, but how to evaluate the combination of the four attributes is interesting to explore. The university and department need to confirm that the application of student thesis examination is the right option.

In this article, we are interested in analyzing the student thesis examination based on private data university—the objectives of this research to analyze the data mechanism. Besides the analyzing mechanism, this research also develops a web application using python programming. It makes simple when the other data come from another university or department and need to analyze to get the association pattern.

II. Method

FP-Growth algorithm was introduced by Han [28]. This algorithm is the improvement of Apriori Algorithms [29] that generate frequent patterns without candidate generation. The association rules mining [30] analyzes frequent patterns that consist of items and finds the association between X and Y, where X and Y are itemsets. The methodology in this research follows Figure 1.
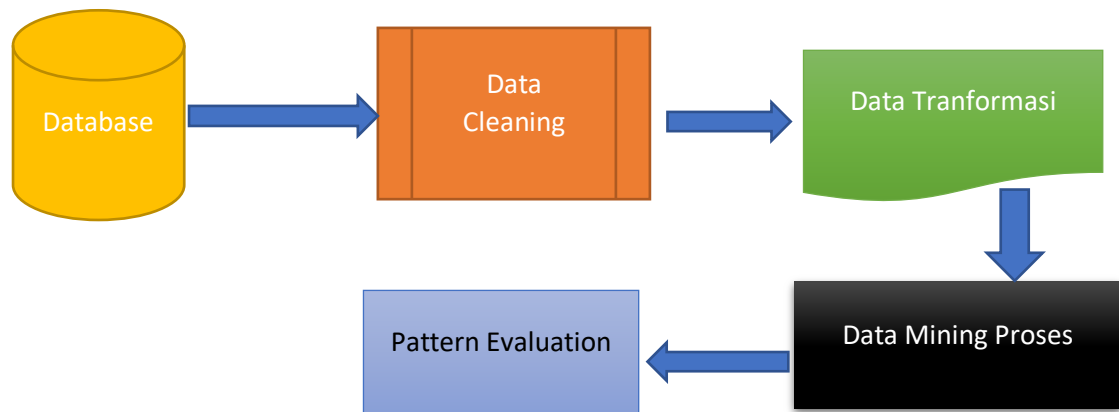

Fig. 1.Methodology to analyze association rules

Starting the development process, previously prepared an e-commerce platform consisting of 2 pages. The first page is the front of the e-commerce platform, and the second is the product detail page. The system scenario that is created assumes that the user explores the system by seeing the products on the front page, continuing to the product detail page. The product item data used is sample data taken from an e-commerce site in Indonesia. The data taken are product names, product prices, pictures, and product descriptions.

1. Data selection. Select the variable based on the independent variable of this research. The variables are research topics, supervisors, and examiners.

2. Data cleaning. Clean the data from noise data. The noise data are the data with null value or irrelevant value, or invalid value. The cleaning mechanism is by removing the noise data.

3. Data load. This research develops a web application using python programming. The Dataset has been prepared then uploaded into the web application to start analyzing.

4. Data transformation. Because of the data in variable research topics, supervisors and examiners can duplicate value. The data need to transform. Hashing technique was chosen in this transformation step.

5. Calculate Frequent Itemsets. Calculate frequent is to find the itemset that frequent of the Dataset. Then determine the minimum support values to eliminate the infrequent itemset.

6. FP-Growth Algorithms. The FP-Growth Algorithms steps consist of several tasks, i.e., building FP-Tree, generate a conditional pattern base, construct conditional FP-Tree, and generate frequent patterns. After frequent patterns generate, then calculate the confidence value of the frequent patterns. The analyzer needs to determine the value of minimum confidence. Then the web application can remove the frequent patterns less than minimum confidence values.

7. Analyze Association Rules. This step is to confirm the validation of the result. If several patterns are not matched with real insight. Then can decide that the historical data have a mismatch

III. Results and Discussion

A. Dataset

The earlier step of this research is to collect the Dataset from the academic unit. The Dataset contains 448-row data with several variables, i.e., student number, name, department, address, data registration, and so on. Table 1 shows the collection of the Dataset. This Dataset used a basis in this research. The next step of this research is preprocessing the Dataset to get a clean dataset processed to the algorithms. The preprocessing step will perform like data selection and data cleansing. This research not only analyses the data but also produces a web application that develops using python programming. The web application also does the preprocessing step but only in data transformation. It is because the data will transform into hash data. We choose the hash data because we need consistent data before the analysis process.

B. Data Selection

The data selection stage is a process for selecting or deciding what variables will be used at the data mining stage. Variables to be used include research topics, supervisors, first examiners, and second examiners. The selection data can be seen, as shown in Table 2. These are related to the objective from the analysis process that we want to understand these variables.

C. Data Cleaning

The cleaning process removes inconsistent data such as noise or invalid data on the supervisor's data, research topics, first examiners, and second examiners. We have done this process for 448 datasets, but unfortunately, there is no data needed to remove.

D. Load Data to Application

A web application has developed using python programming to analyze association rules mining. This python code also uses the pyfpgrowth library to implement the FP-Growth algorithms. After the data cleansing and data selection process, then the Dataset will upload to the web application. The web application will read all of the datasets and store the Dataset into the database.

E. Data Transformation

The data transformation uses function convertToHash in python, as shown in Table 1, to transform the data text into hash data. The hash data only change variable supervisor, first examiner, and second examiner. There are many similarity data in this variable and the fundamental character of association rules every duplicate value will remove. A name can be a supervisor, first examiner, and second examiner. It will duplicate data for association rules, so the solution is by adding a character that can separate the data depending on the variable name. In this case, the supervisor column adds with "pemb" word, the first examiner with "1uji" word, and the second examiner with "2uji" word.

Table 1. Data Transformation Code

No	Code
1	li = { }
2	colname = 'supervisor'
3	li['pemb'] = convertToHash(x, colname)
4	for i,j in li['pemb'].items():
5	x[colname].replace([j]," 'pemb'"+i,inplace=True)
6	colname = 'first examiner'
7	li['1uji'] =convertToHash(x, colname)
8	for i,j in li['1uji'].items():
9	x[colname].replace([j]," '1uji'"+i,inplace=True)
10	colname = 'second examiner'
11	li['2uji'] = convertToHash(x, colname)
12	for i,j in li['2uji'].items():
13	x[colname].replace([j]," '2uji'"+i,inplace=True)

The result of the data transformation process shows in Figure 2. The column supervisor, first examiner, and second examiner have success transform into hash data. By this transformation, all three variables are not duplicate and ready to analyze.
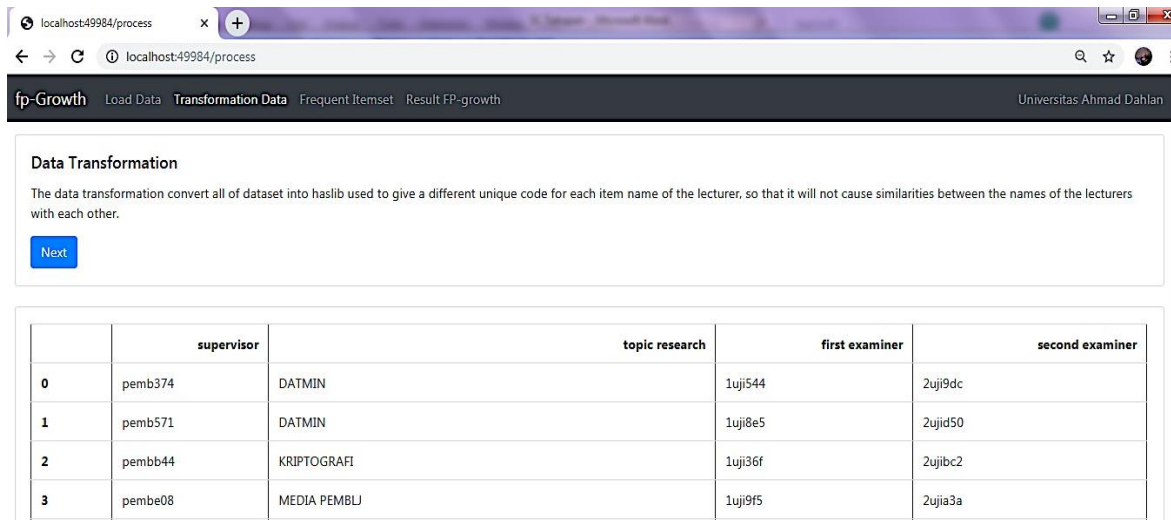
Fig. 2.Data Transformation Step

F. Frequent Itemsets

The first step in the FP-Growth algorithm is to find frequent itemset. This frequent itemset value is needed to decide which item will be used in analyse association rules. The item infrequent will remove, as shown in Figure 3.
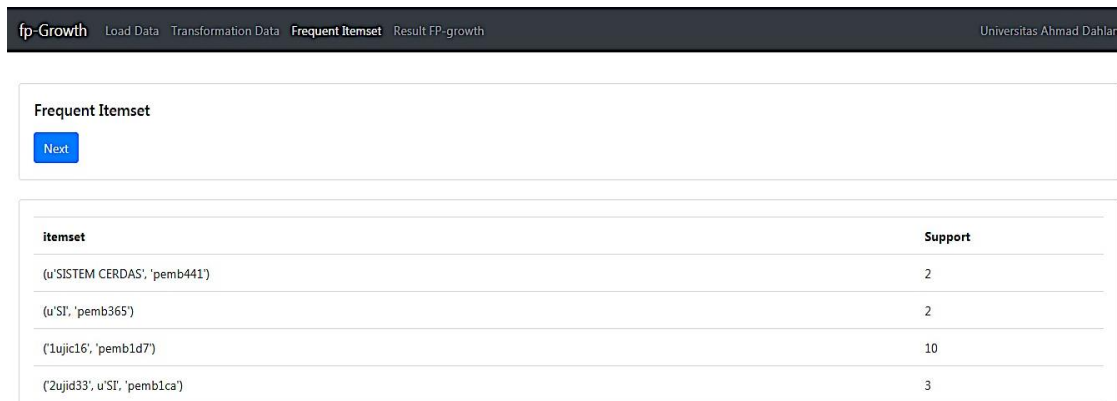

Fig. 3.Result of Frequent Itemset

G. Result of FP-Growth Algorithms

The next step in the FP-Growth algorithm is to determine minimum support and minimum confidence value [29]. In this research, we determine minimal support (min_supp) equal to 4 and minimum confidence (min_conf) equivalent to 100%. The infrequent itemset will remove, and the frequent itemset will re-arrange to build an FP-Tree. Figure 4 shows the illustration of the sample of the FP-Tree in this research. After building FP-Tree, the FP-Growth algorithms need to generate a conditional pattern base, as shown in Table 2.
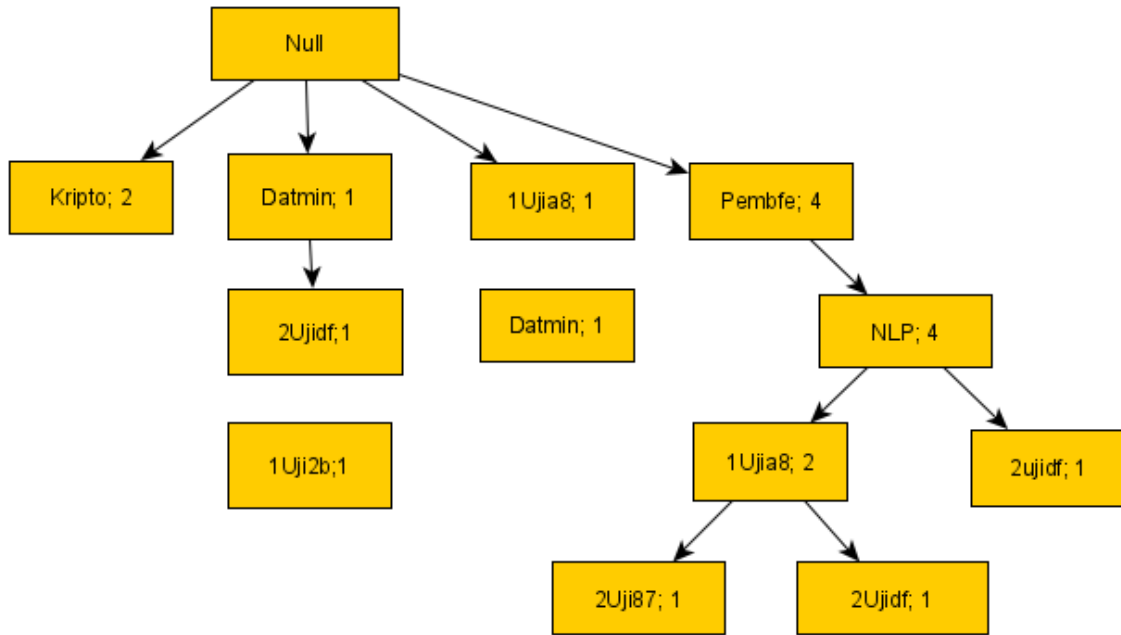
Fig. 4.Sample of FP-Tree

The conditional pattern base is decided based on the tree that has minimum support. The conditional pattern base contains a combination between the item and number of the count from the tree combination. Table 2 is a sample of conditional pattern base in this research.

Table 2. Sample 0f Conditional Pattern Base

suffix	Conditional pattern base
1Ujia8	{(Pembfe,NLP:4)}
2Ujidf	{(Pembfe,NLP:2), (Pembfe , NLP ,1Ujia8:2), (DATMIN:1)}
2Uji87	{(Pembfe,NLP,1Ujia8:4)}

The next step is to generate a conditional FP-Tree. At this stage, to find a conditional FP-Tree, by adding up the existing support count, each item that has a larger number of support counts is equal to the minimum support count that will be generated with a conditional FP-Tree, as shown in Table 3.

Table 3. Sample of Conditional FP-Tree

suffix	Conditional FP-tree
1Ujia8	{(Pembfe ,NLP:4)}
2Ujidf	{(1Ujia8:2), (Pembfe ,NLP:4)}
2Uji87	{(Pembfe , NLP ,1Ujia8:4)}

The conditional FP-Tree will generate Frequent Patterns. This stage is to find a single path and then combine it with items in the FP-Tree conditional. Table 4 shows a sample of Frequent Patterns.

Table 4. Sample of Frequent Pattern

Item	Frequent Patterns Generate
1Ujia8	{(1Ujia8:), (1Ujia8, NLP:3), (1Ujia8,Pmebfe:3), (1Ujia8,NLP,Pembfe:3)}
2Ujidf	{(2Ujidf:), (1Ujia8, 2Ujidf:2), (1Ujia8,2Ujidf,NLP:2), (1Ujia8, 2Ujidf, Pembfe:2), (2Ujidf,NLP:2), (2Ujidf,1Ujia8,NLP:2), (2Ujidf,NLP,1Ujia8:2)(1Ujia8,2Ujidf,Pembfe ,NLP:2)}
2Uji87	{(2Uji87:), (2Uji87,NLP:2), (2Uji,Pembfe:2), (2Uji87,NLP,Pembfe:2}

After frequent patterns are generated, the web application will re-transform the supervisor, first examiner, and second examiner into the real name. The result of the re-transform data shows in Figure 5. Figure 5 also calculates the value of confidence. The values of confidence calculate with Formula 1 [31].

$$Confidence \ (A \Rightarrow B) = \frac{Support\_count(A \cup B)}{support\_count(A)} \qquad (1)$$

The value of minimal confidence in this research is 100%. It is because only the maximal pattern that wants to analyze the result.

H. Analyze Association Rules

The complete result of Figure 5 shows in Table 5. Table 5 is used to analyse step. The focus of the investigation is to count the appropriate pattern if it combines with the background research and topic research of the student. The validation process has done with confirmation to the data background research and data research member. The result of validation found ten patterns (22.22%) is not appropriate. It means the pattern based on historical data is not by the data group research member. It can be due to several reasons FP-growth performance experience a drop when the FP-tree is very dense. The pattern result can be the basis to improve the combination of thesis exam activity in the future.
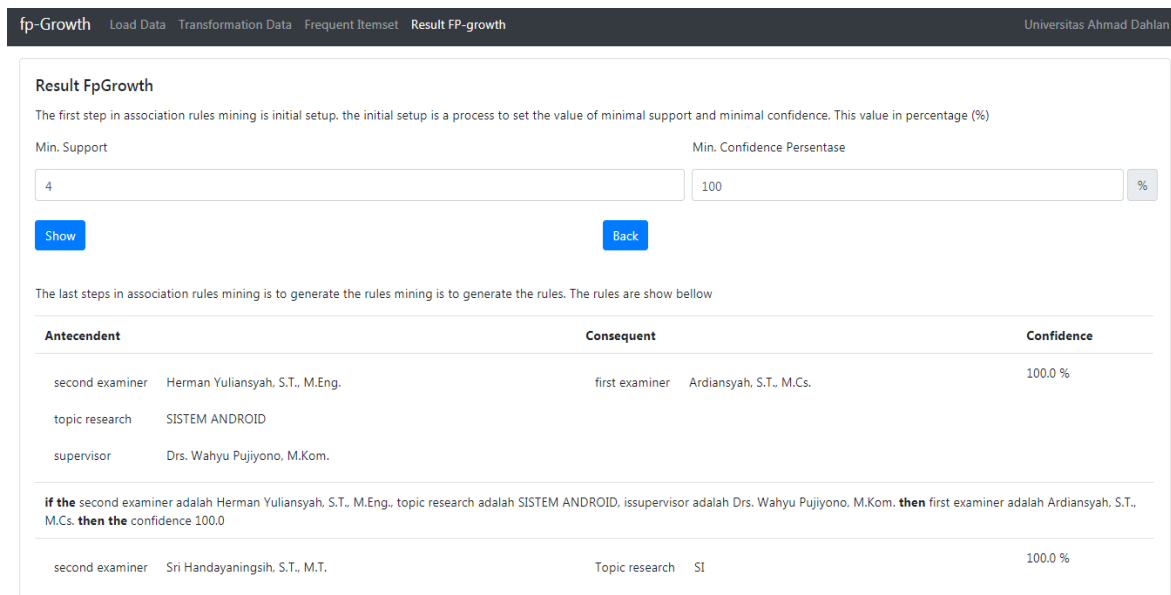

Fig. 5. Minimum Support and Minimum Confidence for FP-Growth Algorithms Results

Table 5. Validation of Asociation Rules Result

No	Association Rules	Validation
1	If the second examiner is Herman Yuliansyah, S.T., M.Eng., topic research is Android Systems, the supervisor is Drs. Wahyu Pujiyono, M.Kom. Then the first examiner is Ardiansyah, S.T., M.Cs.	Appropriate
2	If the second examiner is Sri Handayaningsih, S.T., M.T., supervisor is Eko Aribowo, S.T., M.Kom. Then topic research is Information System	Appropriate
3	If the second examiner is Ir. Ardi Pujiyanta, M.T., topic research is Intelligence Systems, the supervisor is Murinto, S.Si., M.Kom. Then the first examiner is Kartika Firdausy, S.T., M.T.	Appropriate
4	If the topic research is Android Systems, the supervisor is Drs. Wahyu Pujiyono, M.Kom. Then the first examiner is Ardiansyah, S.T., M.Cs	Appropriate
5	If the first examiner is Drs. Wahyu Pujiyono, M.Kom., the second examiner is Murinto, S.Si., M.Kom. Then topic research is Multimedia	Not Appropriate
6	If the first examiner is Drs. Wahyu Pujiyono, M.Kom., the second examiner is Murinto, S.Si., M.Kom., the supervisor is Fiftin Noviyanto, S.T., M.Cs. Then topic research is Multimedia	Not Appropriate
7	If the first examiner is Kartika Firdausy, S.T., M.T., the second examiner is Ir. Ardi Pujiyanta, M.T. Then the supervisor is Murinto, S.Si., M.Kom.	Appropriate

No	Association Rules	Validation
8	If the first examiner is Eko Aribowo, S.T., M.Kom., the second examiner is Dewi Soyusiawaty, S.T., M.T. Then the supervisor is Sri Winiarti, S.T., M.Cs.	Not Appropriate
9	If the first examiner is Ardiansyah, S.T., M.Cs., the supervisor is Eko Aribowo, S.T., M.Kom. Then topic research is Information System	Appropriate
10	If the topic research is Intelligence Systems, the supervisor is Ir. Ardi Pujiyanta, M.T. Then the second examiner is Sri Winiarti, S.T., M.Cs.	Appropriate
11	If the first examiner is Murinto, S.Si., M.Kom., the supervisor is Dr. Abdul Fadlil, M.T. Then topic research is Intelligence Systems	Appropriate
12	If the second examiner is Ir. Ardi Pujiyanta, M.T., the topic research is Intelligence Systems, Then the supervisor is Murinto, S.Si., M.Kom.	Appropriate
13	If the second examiner is Ali Tarmuji, S.T., M.Cs., topic research is Geography Information System, the supervisor is Eko Aribowo, S.T., M.Kom. Then the first examiner is Drs. Tedy Setiadi, M.T.	Appropriate
14	If the first examiner is Drs. Tedy Setiadi, M.T., topic research is Geography Information System, the supervisor is Eko Aribowo, S.T., M.Kom. Then the second examiner is Ali Tarmuji, S.T., M.Cs.	Appropriate
15	If the second examiner is Ali Tarmuji, S.T., M.Cs., the supervisor is Sri Handayaningsih, S.T., M.T., M.T. Then the topic research is Information Systems	Appropriate
16	If the first examiner is Kartika Firdausy, S.T., M.T., the topic research is Intelligence System Then the supervisor is Murinto, S.Si., M.Kom.	Appropriate
17	If the second examiner is Herman Yuliansyah, S.T., M.Eng., supervisor is Drs. Wahyu Pujiyono, M.Kom. Then topic research is Android Systems	Appropriate
18	If the first examiner is Drs. Wahyu Pujiyono, M.Kom., the supervisor is Sri Handayaningsih, S.T., M.T. Then the topic research is Information System	Appropriate
19	If the first examiner is Drs. Tedy Setiadi, M.T., the second examiner is Rusydi Umar, S.T., M.T., Ph.D. Then the supervisor is Fiftin Noviyanto, S.T., M.Cs.	Appropriate
20	If the supervisor is Dr. Abdul Fadlil, M.T. Then Topic research is Intelligence System	Appropriate
21	If the first examiner is Dr. Abdul Fadlil, M.T., the second examiner is Sri Winiarti, S.T., M.Cs., topic research is Intelligence System, Then the supervisor is Ir. Ardi Pujiyanta, M.T.	Appropriate
22	If the first examiner is Murinto, S.Si., M.Kom., the topic research is Data Mining. Then the supervisor is Drs. Tedy Setiadi, M.T.	Not Appropriate
23	If the topic research is Information System, the supervisor is Sri Handayaningsih, S.T., M.T., M.T. Then the second examiner is Ali Tarmuji, S.T., M.Cs.	Appropriate
24	If the first examiner is Kartika Firdausy, S.T., M.T., the second examiner is Ir. Ardi Pujiyanta, M.T., the topic research is Intelligence System. Then the supervisor is Murinto, S.Si., M.Kom.	Appropriate
25	If the second examiner is Murinto, S.Si., M.Kom., the supervisor is Fiftin Noviyanto, S.T., M.Cs. Then topic research is Multimedia	Appropriate
26	If the first examiner is Dr. Abdul Fadlil, M.T., the supervisor is Sri Winiarti, S.T., M.Cs. Then topic research is Intelligence System	Appropriate
27	If the second examiner is Murinto, S.Si., M.Kom., the topic research is Intelligence System. Then the supervisor is Sri Winiarti, S.T., M.Cs.	Appropriate
28	If the supervisor is Sri Handayaningsih, S.T., M.T., M.T. Then the second examiner is Ali Tarmuji, S.T., M.Cs., the topic research is Information System	Appropriate
29	If the second examiner is Murinto, S.Si., M.Kom., the supervisor is Drs. Wahyu Pujiyono, M.Kom. Then topic research is Multimedia	Not Appropriate
30	If the first examiner is Drs. Wahyu Pujiyono, M.Kom., the second examiner is Ali Tarmuji, S.T., M.Cs. Then topic research is Information System	Appropriate
31	If the second examiner is Sri Handayaningsih, S.T., M.T., M.T. Then the first examiner is Drs. Tedy Setiadi, M.T.	Appropriate
32	If the first examiner is Drs. Wahyu Pujiyono, M.Kom., the second examiner is Ali Tarmuji, S.T., M.Cs., the supervisor is Sri Handayaningsih, S.T.,M.T. Then topic research is Information System	Appropriate
33	If the first examiner is Drs. Tedy Setiadi, M.T., topic research is Information System, the supervisor is Sri Handayaningsih, S.T., M.T., M.T. Then the second examiner is Ali Tarmuji, S.T., M.Cs.	Appropriate
34	If the topic research is Geography Information System, the supervisor is Eko Aribowo, S.T., M.Kom. Then first examiner is Drs. Tedy Setiadi, M.T.	Appropriate
35	If the first examiner is Murinto, S.Si., M.Kom., the supervisor is Drs. Wahyu Pujiyono, M.Kom. Then topic research is Multimedia	Not Appropriate
36	If the first examiner is Kartika Firdausy, S.T., M.T., the second examiner is Sri Winiarti, S.T., M.Cs. Then the supervisor is Murinto, S.Si., M.Kom.	Appropriate
37	If the first examiner is Dr. Abdul Fadlil, M.T., the second examiner is Sri Winiarti, S.T., M.Cs., the supervisor is Ir. Ardi Pujiyanta, M.T. Then topic research is Intelligence System	Appropriate

No	Association Rules	Validation
38	If the first examiner is Drs. Tedy Setiadi, M.T., the second examiner is Ali Tarmuji, S.T., M.Cs., the supervisor is Sri Handayaningsih, S.T., M.T., M.T. Then topic is Information System	Appropriate
39	If the first examiner is Murinto, S.Si., M.Kom., the second examiner is Fiftin Noviyanto, S.T., M.Cs., the supervisor is Drs. Wahyu Pujiyono, M.Kom. Then topic research is Multimedia	Not Appropriate
40	If the first examiner is Drs. Wahyu Pujiyono, M.Kom., the supervisor is Fiftin Noviyanto, S.T., M.Cs. Then the second examiner is Murinto, S.Si., M.Kom.	Not Appropriate
41	If the first examiner is Drs. Wahyu Pujiyono, M.Kom., the topic research is Multimedia, the supervisor is Fiftin Noviyanto, S.T., M.Cs. Then the second examiner is Murinto, S.Si., M.Kom.	Not Appropriate
42	If the topic research is Intelligence System, the supervisor is Dr. Abdul Fadlil, M.T. Then the first examiner is Murinto, S.Si., M.Kom.	Appropriate
43	If the first examiner is Drs. Tedy Setiadi, M.T., the supervisor is Sri Handayaningsih, S.T., M.T., M.T. Then Topic research is Information System	Appropriate
44	If the first examiner is Murinto, S.Si., M.Kom., the second examiner is Fiftin Noviyanto, S.T., M.Cs., the topic research is Multimedia. Then the supervisor is Drs. Wahyu Pujiyono, M.Kom.	Not Appropriate
45	If the first examiner is Ardiansyah, S.T., M.Cs., the second examiner is Herman Yuliansyah, S.T., M.Eng., the supervisor is Drs. Wahyu Pujiyono, M.Kom. Then topic research is Android Systems	Appropriate

IV. Conclusion

Based on the result and discussion section, searching the Frequent Itemset of the decision tree uses the FP-Growth algorithm that works very well in doing the Frequent Itemset with the FP-Tree building process by generating rules from existing datasets. The results found 45 patterns of association rules with a minimum support value is four, and minimum confidence is 100%. This pattern has been validated with scientific group data, found 77.78% appropriate pattern, and 22.22% is not appropriate.

References

[1] J. Silva, N. Varela, L. A. Borrero López, and R. H. Rojas Millán, "Association Rules Extraction for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm," *Procedia Comput. Sci.*, vol. 151, pp. 1207–1212, 2019.

[2] H. Yuliansyah, Hafsah, I. Arfiani, and R. Umar, "Discovering Meaningful Pattern of Undergraduate Students Data using Association Rules Mining," in *2019 Ahmad Dahlan International Conference Series on Engineering and Science (ADICS-ES 2019)*, 2019, pp. 13–17.

[3] K. . Lakshmi and G. Vadivu, "Extracting Association Rules from Medical Health Records Using Multi-Criteria Decision Analysis," *Procedia Comput. Sci.*, vol. 115, pp. 290–295, 2017.

[4] S. Winiarti, H. Yuliansyah, and A. A. Purnama, "Identification of Toddlers' Nutritional Status using Data Mining Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 164–169, 2018.

[5] V. Ivančević, I. Tušek, J. Tušek, M. Knežević, S. Elheshk, and I. Luković, "Using association rule mining to identify risk factors for early childhood caries," *Comput. Methods Programs Biomed.*, vol. 122, no. 2, pp. 175–181, Nov. 2015.

[6] S. Altuntas, T. Dereli, and A. Kusiak, "Analysis of patent documents with weighted association rules," *Technol. Forecast. Soc. Change*, vol. 92, pp. 249–262, Mar. 2015.

[7] Z. A. Othman, N. Ismail, and M. T. Latif, "Association rules of temperature towards high and low ozone in Putrajaya," in *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*, 2017, pp. 1–5.

[8] Z. A. Othman, N. Ismail, and M. T. Latif, "Association Pattern of NO2 and NMHC towards High

Ozone Concentration in Klang," in *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*, 2017, no. 2, pp. 1–6.

[9]   I. Riadi, S. Winiarti, and H. Yuliansyah, "Development and Evaluation of Android Based Notification System to Determine Patient' s Medicine for Pharmaceutical Clinic," *2017 4th Int. Conf. Electr. Eng. Comput. Sci. Informatics*, no. September, pp. 19–21, 2017.

[10]   S. K. Verma, R. S. Thakur, and S. Jaloree, "Fuzzy association rule mining based model to predict students' performance," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 4, pp. 2223–2231, 2017.

[11]   N. Chaimongkol and P. Meesad, "Association Rule Mining for Specific New Course," *J. Inf. Sci. Technol.*, vol. 1, no. 1, pp. 15–22, Jun. 2010.

[12]   P. Paruechanon and W. Sriurai, "Applying association rule to risk analysis for student-dropout in Information Technology Program," *J. Sci. Sci. Educ.*, vol. 1, no. 2, pp. 123–133, 2018.

[13]   B. Mahatthanachai, H. Ninsonti, and N. Tantranont, "A Study of Factors Influency Student Dropout Rate Using Data Mining," *Golden Teak Humanit. Soc. Sci. J.*, vol. 22, no. 4, pp. 46–55, 2016.

[14]   B. Pérez, C. Castellanos, and D. Correal, "Predicting student drop-out rates using data mining techniques: A case study," in *Communications in Computer and Information Science*, 2018.

[15]   J. Chareonrat, "Student drop out factor analysis and trend prediction using decision tree," *Suranaree J. Sci. Technol.*, 2016.

[16]   B. Perez, C. Castellanos, and D. Correal, "Applying Data Mining Techniques to Predict Student Dropout: A Case Study," in *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence, ColCACI 2018 - Proceedings*, 2018.

[17]   K. Sya'iyah, H. Yuliansyah, and I. Arfiani, "Clustering Student Data Based On K-Means Algorithms," *Int. J. Sci. Technol. Res.*, vol. 8, no. 8, pp. 1014–1018, 2019.

[18]   P. Belsis, I. Chalaris, M. Chalaris, C. Skourlas, and A. Tsolakidis, "The Analysis of the Length of Studies in Higher Education based on Clustering and the Extraction of Association Rules," *Procedia - Soc. Behav. Sci.*, vol. 147, pp. 567–575, 2014.

[19]   M. R. Anderson-Rowland *et al.*, "Mining students data to analyze learning behavior: A case study," *Indones. J. Electr. Eng. Comput. Sci.*, 2018.

[20]   W. Jie, H. Y. Lv, C. Biao, and Z. Yuan, "Application of educational data mining on analysis of students' online learning behavior," in *2017 2nd International Conference on Image, Vision and Computing, ICIVC 2017*, 2017.

[21]   S. H. Mankad, "Predicting learning behaviour of students: Strategies for making the course journey interesting," in *Proceedings of the 10th International Conference on Intelligent Systems and Control, ISCO 2016*, 2016.

[22]   R. Sumitha, E. S. Vinothkumar, and P. Scholar, "Prediction of Students Outcome Using Data Mining Techniques," *Int. J. Sci. Eng. Appl. Sci.*, 2016.

[23]   M. Wati, W. Indrawan, J. A. Widians, and N. Puspitasari, "Data mining for predicting students' learning result," in *Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017*, 2018.

[24]   B. Guo, R. Zhang, G. Xu, C. Shi, and L. Yang, "Predicting Students Performance in Educational Data Mining," in *Proceedings - 2015 International Symposium on Educational Technology, ISET 2015*, 2016.

[25]   R. Bansal, A. Mishra, and S. N. Singh, "Mining of educational data for analysing students' overall performance," in *Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering*, 2017.

[26]   R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, 2017.

[27]   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. Adv. Comput. Res.*, vol. 6, no. 27, pp. 215–221, Oct. 2016.

[28]   J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," *ACM sigmod Rec.*, vol. 29, no. 2, 2000.

[29]   R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Rec.*, vol. 22, no. 2, pp. 207–216, 1993.

[30]   R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," *J. Comput. Sci. Technol.*, vol. 15, no. 6, pp. 487–499, 1994.

[31]   J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. USA: Morgan Kaufmann, 2012.