

Determination of Student Subjects in Higher Education Using Hybrid Data Mining Method with the K-Means Algorithm and FP Growth

Larissa Navia Rani^{a,1,*}, Sarjon Defit^{b,2}, L. J. Muhammad^{c,3}

^{a,b} Universitas Putra Indonesia YPTK Padang, Jl. Raya Lubuk Begalung, Padang, 25221, Indonesia

Federal University, Kashere, Gombe State, Nigeria

¹ larissa_navia_rani@upiypk.ac.id*; ² sarjond@yahoo.co.uk

* corresponding author

ARTICLE INFO

ABSTRACT

Article history:

Received 15-04-2021

Revised 03-06-2021

Accepted 16-07-2021

Keywords:

Hybrid Data Mining

K-Means Algorithm

FP Growth Algorithm

Higher Education

The large number of courses offered in an educational institution raises new problems related to the selection of specialization courses. Students experience difficulties and confusion in determining the course to be taken when compiling the study plan card. The purpose of this study was to cluster student value data. Then the values that have been grouped are seen in the pattern (pattern) of the appearance of the data based on the values they got previously so that students can later use the results of the patterning as a guideline for taking what skill courses in the next semester. The method used in this research is the K-Means and FP-Growth methods. The results of this rule can provide input to students or academic supervisors when compiling student study plan cards. Lecturers and students can analyze the right specialization subject by following the pattern given. This study produces a pattern that shows that the specialization course with the theme of business information systems is more followed by students than the other 2 themes.

Copyright © 2021 International Journal of Artificial Intelligence Research.
All rights reserved.

I. Introduction

Higher Education is one of the educational institutions in Indonesia. Universities carry out education at the highest level. Students who attend tertiary education must have completed education at the previous level, namely secondary education with the name high school (SMA). Higher education also has four different levels of education in it starting from the first is diploma 3 (D3) to undergraduate (S1) or undergraduate, then strata 2 (S2) or master, and the last is strata 3 (S3) or doctoral. Educators in higher education are given the name of the lecturer while the students are given the name of the student [1], [2]. The relationship between lecturers and students is the same as the relationship between teachers and students (students) which applies at the secondary and basic education levels. In carrying out the educational process, higher education institutions use the credit system (semester credit units). SKS is a unit load of courses applicable in higher education that must be taken by every student. The courses that are presented in the form of credits must be completed by students every semester. For undergraduate students, the total credit load that must be completed (completed) by students during their education is at least 144 credits. The total credits represent the number of courses that students take while carrying out lectures. Each course has a credit rating of 2 credits, 3 credits, and 4 credits, respectively. So during their undergraduate education students must complete approximately 30-45 courses [3], [4].

Subjects are lecture material that must be studied and completed by students while studying in college. Each student's course differs from one study program to another. Subjects depend on the field of study or expertise has taken by students. The subjects are taken by students in a study program also consist of five types, namely Personality Development Courses (MPK), Science and Skills Courses (MKK), Work Skills Courses (MKB), Rich Behavior Courses (MPB), and Subjects Living Together

Lecture (MBB). Subjects that describe science, provide knowledge and skills to students in their study program are expertise courses [3], [5]. The education system has a lot of data that can be retrieved and processed to produce new knowledge. One that includes educational data is a subject. The large number of courses offered in an educational institution raises new problems related to the selection of specialization courses. Students experience difficulties and confusion in determining the course to be taken when compiling the study plan card. Errors in choosing elective courses will affect the student's grades and will also affect the overall student performance.

Higher education provides freedom to students in taking expertise courses. The campus provides many choices of expertise courses to students that can be taken. Not all skill courses must be taken by students. It depends on the student who wants to take which course, based on their interests, curiosities, and talents. These expertise courses are usually only provided in the final semesters of lectures such as semesters 5, 6, 7, or 8. The goal is that students first complete general courses and new compulsory courses so that they can then choose expertise courses. Also, the aim is to provide students with a strong foundation of knowledge first, then expertise as development. In fact, in the student field, many are hesitant to take what kind of expertise course. Students are often confused about what courses they are suitable for. Many students are afraid that later taking this course will be difficult to pass, it is difficult to get good grades, and so on. Many students feel insecure about the abilities they have that they will be able to complete these courses. In general, students only follow other friends in taking courses so that later they can easily complete the course. If students already know very well their abilities in the previous course, they will feel confident taking this course.

The purpose of this research is to do two things, namely grouping (clustering) student value data. The clustering is to calculate how many students have good grades and how many students have poor grades and grouping these scores in any subject. Then the second is that the values that have been grouped are seen in the pattern (pattern) of the appearance of the data based on the values they got previously so that students can later use the results of the patterning as a guideline for taking what skill courses in the next semester. Many ways can be used by educational institutions to perform clustering and forecasting (patterns). To carry out the clustering process, researchers used the K-Means method [6], [7] and the results of the clustering process using the K-Means method [8], [9] were used for pattern making, for this reason, the researcher using the FP-Growth method [10], [11]. The K-Means method has been widely used to perform the data clustering process because the process is not too complicated and the results are accurate [12], [13]. While the FP-Growth method has also been widely used to process data patterns because the process is fast and the results are precise [14], [15].

One of the large and well-known private universities in Indonesia in the province of West Sumatra is Putra Indonesia University YPTK Padang. This university has a student population of approximately 15,000 active students. With so many students, of course, they have problems taking expertise courses. Therefore this research was carried out. This paper is composed of four parts, namely an introduction, then the research methodology, then the results, and a final discussion of the conclusions.

Previous research that has been conducted and is relevant to this research is by Sufajar Butsianto et al in 2020 who came from Pelita Bangsa University, Bekasi, West Java [16]. In this study, the results obtained are grouping (cluster) for data similarity using the k-means algorithm, so that data with the same characteristics will be in one cluster. The attributes used are brand and sales. Clusters formed after the K-Means Clustering process were divided into three clusters, namely Cluster 0 with a percentage of 235 members with a percentage of 26% categorized as Laris, Cluster 1 with a percentage of 604 with a percentage of 67% categorized as Less in demand, and Cluster 2 the number of 61 members with a percentage of 7% categorized as Best Selling, from the clustering process above, it can be obtained the DBI (Davies Bouldin Index) validation with a value of 0.34. The drawback of this study is that it only uses one method, namely the K-Means method, so the results are less accurate.

The second research conducted by Kristoko Dwi Hartomo et al in 2020 came from Satya Wacana Christian University, Salatiga, Central Java [17]. The FP-Growth Algorithm is used to evaluate the trend of customer behavior in this analysis, so that it can be used for decision making in the preparation of goods and the Triple Exponential Smoothing algorithm, which is a prediction algorithm, is used to control stock items. The tea and sugar association rules had the highest lift ratio value of 6.131 in this analysis, and with the Triple Exponential Smoothing algorithm, the expected results in January 2018

were 131.141 kg with an 88.3 percent MAPE accuracy. The weakness in this study is that the data used as training data is less than 10,000 transactions so that the calculation of predictions made is less accurate.

II. Methodology

A. Research Framework

When it comes to science, of course, you must follow the rules or a structured and systematic research framework. The research framework, can be seen in fig. 1.

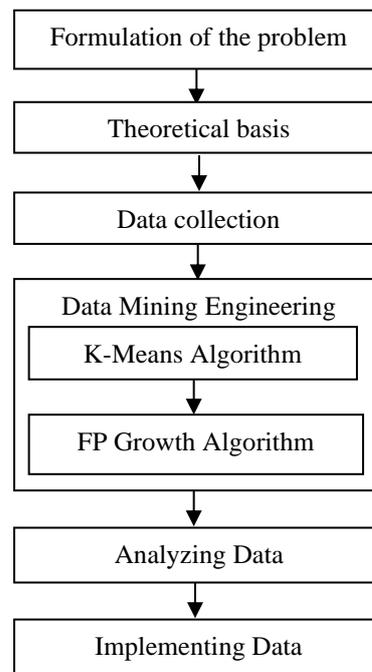


Fig. 1. Research framework

B. Research Framework Details

Problem determination is the process of determining what are the main and important problems that occur for which a solution should be found in this research. Theory study is the process of finding or examining previous sciences and research relevant to solving predetermined problems. Data collection is the process of collecting field data that will be used in problem-solving. In this study, the data were sourced from the Academic Bureau (PDE and Sisfo) of Universitas Putra Indonesia YPTK Padang. The number of lines of testing data to be tested that has been collected is 16 lines of data or 16 in other words, subjects. To do data clustering using the K-Means method, the researcher used the data from 16 subjects. Then in making a pattern using the FP-Growth method also uses the same data. The data collected consists of 15 data attributes as in table 1 below:

Table 1. Data Attributes

No	Data Attributes	Data Types
1	Abyan Falmi	Real
2	Andre Ignatius Nyonathan	Real
3	Christopher	Real
4	Desi Warnitati	Real
5	Dicky Fernando	Real
6	Dinda Lestari	Real
7	Ellsa Firdaus	Real
8	Fadhillah Afif Septian	Real
9	Fajri Karim	Real
10	Abdul Zahki	Real
11	Iin Dara Sinta	Real

12	Indah Dwi Putri Rahmadani	Real
13	Mira Lestari	Real
14	Mu'tasim Billah Syani	Real
15	Nofita Sari	Real

At this stage, we use a Wiener transformation approach. The Wiener Transformation calculation approach is assisted by calculating the mean (μ) and variance (σ^2) Means and variants are calculated from the input x, that is, the raw data is cleaned. This approach works by converting data in the non-numeric form to numeric, which aims to eliminate the missing value from the data to be used. The calculation of the mean and variant is as in the following equation:

Mean and Variance value calculation formula:

$$\mu = \frac{1}{pq} \sum_{n_1 n_2 \in n} X(n_1, n_2) \quad (1)$$

$$\sigma^2 = \frac{1}{pq} \sum_{n_1 n_2 \in n} (X^2(n_1, n_2) - \mu) \quad (2)$$

Where: μ = average values per input field, σ^2 = variant, p = input row, q = input coloum, n_1 = data to-1, n_2 = data to-n, X = matrix input 2 dimensions, X^2 = matrix input quadrant 2 dimension

To get the value of $y(n_1, n_2)$ as the output of the Wiener Transformation method, it can be seen in the following equation:

$$v^2 = \frac{\sigma^2}{\sum \sigma^2} \quad (3)$$

Value calculation formula of $y = (n_1, n_2)$:

$$Y(n_1, n_2) = \mu + \frac{\sigma^2 - v^2}{\sigma^2} (X(n_1, n_2) - \mu) \quad (3)$$

Where: σ^2 = variant, v^2 = variant average, n_1 = data to-1, n_2 = data to-n, X^2 = matrix input quadrant 2 dimension, μ = average input column, Y = matrix output 2 dimension.

The preprocessing stages to eliminate missing values with Wiener Transformation are as follows:

- The mean value (μ) was calculated from the student's KRS data into a $1 \times q$ (student ID) matrix from the value conversion data into numbers. Here is the original data we use at this stage.
- The value of variance (σ^2) was calculated from the data into the $1 \times q$ matrix (student ID).
- The mean of variance (v^2) is calculated into the scalar matrix.
- $y(n_1, n_2)$ is the result of calculating the mean and variance that has been obtained in the form of a matrix p (subject) $\times q$ (student ID).

1. Data Processing Using the K-Means Algorithm

The data that the authors have collected in the field are processed using the K-Means method.

The following are the steps for the K-Means method, namely [18], [19]:

1. Determining the number of data clusters to be created, the number of data clusters is called the k value
2. Determine the mean (centroid value) randomly for each predetermined group.
3. Determine the nearest cluster center on each data row with the centroid value, to determine this value using the formula [20]:

$$d_{Euclidean}(x, y) = \sqrt{\sum(x_i - y_i)^2} \quad (4)$$

Where: $d_{Euclidean}(x, y)$ = distance value for each data row with centroid value, $x = x_1, x_2, x_3, etc$, $y = y_1, y_2, y_3, dst$

4. Determine the closest group (cluster) for each row of data by comparing the value of the closest distance that has been obtained in the previous process than updating the center value of the group using the formula:

$$Cluster\ Center = \sum \frac{a_i}{n} \quad (5)$$

Where: $Cluster\ Center$ = the cluster center value, a_i = value on each cluster, n = number of clusters.

5. Repeating steps 3 to step 5 until there is no transfer of data for each row of data from one group to another.

2. Data Processing Using FP Growth Algorithm

The three key stages of the FP-Growth process are as follows, namely [21]:

1. The stage in which a conditional pattern base is formed
2. The conditional generation stage of the FP-Tree
3. The quest stage for frequently used itemsets

The process of determining all associative rules that meet the minimum criteria for help (minsup) and configuration (minconf) in a database is known as association rule mining. About the stated constraints, the minsup and minconf, these two conditions will be used for interesting association rules. Rule of Association A method for determining relationships between objects in a dataset is known as mining. Begin by searching for frequent itemsets, which are the combinations that appear in an itemset the most frequently and must reach the minsup. In this step, the database will be searched for item combinations that meet the minimum criteria for support values. The following formula can be used to calculate the support value of item A. [22][23]:

$$Support(A) = \frac{\text{number of transaction contain (A)}}{\text{number of transaction}} \quad (6)$$

The help value of the two objects is then calculated using the following formula:

$$Support(AUB) = \frac{\text{number of transaction contain (AUB)}}{\text{number of transaction}} \quad (7)$$

The minimum trust (mincof) requirement can be calculated using the following formula after all frequent items and large itemsets have been obtained:

$$Confidence(B/A) = \frac{\text{number of transaction contain (AUB)}}{\text{number of transaction A}} \quad (8)$$

III. Result and Discussion

A. Results of Data Processing Using the K-Means Algorithm

After grouping using the K-Means method, the results are in the form of courses into several groups or clusters which can be seen in Table 2 below:

Table 2. Clusterization Results

No	Nama	Cluster		
		Cluster 0	Cluster 1	Cluster 2
1	Information Technology Quality Management		V	
2	Information Technology Risk Management		V	
3	Integration of Corporate and Information Applications		V	
4	Information Technology Audit		V	
5	Change Management & IT Projects		V	
6	E-Business	V		
7	Technology Business Concepts	V		
8	Information Systems Project Management	V		
9	Business Process Restructuring	V		
10	Supply Chain Management	V		
11	Customer Relationship Management	V		
12	Automation System			V
13	E-commerce			V
14	Applied Artificial Intelligent			V
15	Geographical Information System			V
16	Enterprise Resource Planning			V

The Rapid Miner application was used to process the above data. The results of processing with the Rapid Miner application are shown in Fig. 2.

Cluster Model

```
Cluster 0: 6 items
Cluster 1: 5 items
Cluster 2: 5 items
Total number of items: 16
```

Fig. 2. Cluster Model

The members of the data record for each predetermined cluster which can be seen in Figure 2 above To get the above results, it takes a middle value or centroid value to produce clusters. Below is the centroid value.

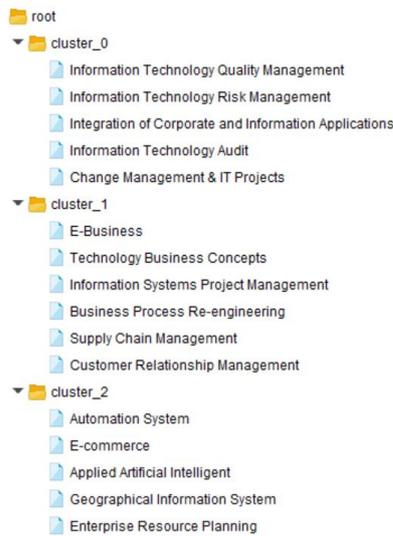


Fig. 3. Existing Cluster Members

Table 3. Centroid Value

Attributes	Cluster 0	Cluster 1	Cluster 2
Abyan Falmi	1.346	2.810	1.353
Andre Ignatius Nyonathan	-15.842	1.050	1.353
Christopher	1.346	2.912	1.353
Desi Warnitati	1.346	1.050	-6.326
Dicky Fernando	1.346	2.709	1.353
Dinda Lestari	1.346	2.811	1.353
Ellsa Firdaus	1.346	2.555	1.353
Fadhillah Afif Septian	1.346	1.991	1.353
Fajri Karim	1.346	2.698	1.353
Abdul Zahki	1.346	1.050	-5.268
Iin Dara Sinta	1.346	2.454	1.353
Indah Dwi Putri Rahmadani	1.346	2.595	1.353
Mira Lestari	1.346	2.912	1.353
Mu'tasim Billah Syani	1.346	2.59	1.353

B. Results of Data Processing Using the FP Growth Method

Look for a collection of frequently used items. Count the number of times each object appears in table 4 below, then sort from the highest to the smallest occurrence frequency. Give ID to each set item to make it easier to carry out the next steps.

Table 4. Ordered Frequent Item Set

ID	Subject	Frequency
6	E-Business	12
9	Business Process Restructuring	12
11	Customer Relationship Management	12
10	Supply Chain Management	11
7	Technology Business Concepts	9
8	Information Systems Project Management	3
12	Automation System	2
13	E-commerce	2
14	Applied Artificial Intelligent	2
15	Geographical Information System	2
16	Enterprise Resource Planning	2
1	Information Technology Quality Management	1
2	Information Technology Risk Management	1
3	Integration of Corporate and Information Applications	1
4	Information Technology Audit	1
5	Change Management & IT Projects	1

After the item priority process has been carried out, the next step is to create an FP-Tree based on the order of priority items. Making the FP-Tree refers to table 4 to see the activity of each sample and then sorted from the largest to the smallest frequency. The following is the FP tree generated at this stage. The results data in the table above are related to the ID attribute data of each. The following is the result of this processing.

Table 5. Converted FP-Tree according to the given ID

Name	Score				
Abyan Falmi	6	9	11	10	7
Andre Ignatius Nyonathan	1	2	3	4	5
Christopher	6	9	11	10	7
Desi Warnitati	12	13	14	15	16
Dicky Fernando	6	9	11	10	7
Dinda Lestari	6	9	11	10	7
Ellsa Firdaus	6	9	11	10	8
Fadhillah Afif Septian	6	9	11	8	8
Fajri Karim	6	9	11	10	7
Abdul Zahki	12	13	14	15	16
In Dara Sinta	6	9	11	10	8
Indah Dwi Putri Rahmadani	6	9	11	10	7
Mira Lestari	6	9	11	10	7
Mu'tasim Billah Syani	6	9	11	10	7
Nofita Sari	6	9	11	10	7

The last stage is looking for the support value and the confidence value according to the specified minimum support value = 75% and confidence = 80%, while the items used are {6,9,11,10,7} → this number is obtained from the frequent item set. The following are the results of using the Rapid Miner Application to process data:

No.	Premises	Conclusion	Support ↓	Confidence	LaPlace
47	Business Process Re-engineering	Customer Relationship Management	0.800	1	1
48	Customer Relationship Management	Business Process Re-engineering	0.800	1	1
49	Business Process Re-engineering	e-Bisnis	0.800	1	1
50	e-Bisnis	Business Process Re-engineering	0.800	1	1
54	Customer Relationship Management	e-Bisnis	0.800	1	1
55	e-Bisnis	Customer Relationship Management	0.800	1	1
63	Business Process Re-engineering	Customer Relationship Management, e-Bisnis	0.800	1	1
64	Customer Relationship Management	Business Process Re-engineering, e-Bisnis	0.800	1	1
65	Business Process Re-engineering, Customer Rel.	e-Bisnis	0.800	1	1
66	e-Bisnis	Business Process Re-engineering, Customer Rel.	0.800	1	1
67	Business Process Re-engineering, e-Bisnis	Customer Relationship Management	0.800	1	1
68	Customer Relationship Management, e-Bisnis	Business Process Re-engineering	0.800	1	1

Fig. 4. Results of Rapid Miner Processing

Figure 4. Results of Rapid Miner Processing To produce the data above, a Rapid Miner Application block design is needed:

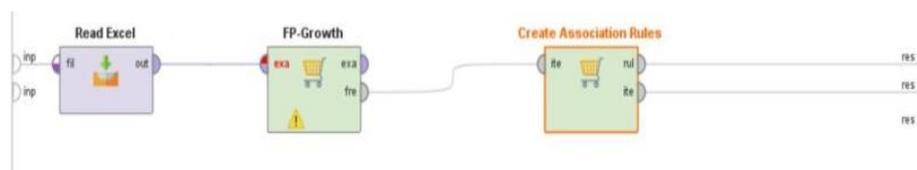


Fig. 5. FP Growth Rapid Miner Block Diagram

In fig. 5 above we can see the Rapid Miner Blog Diagram that is used to do Data Mining with FP-Growth Algorithm. From the left figure is the first block is Read Excell. Read Excell means that system on Rapid Miner reading data from Excell data that be input to the system. The second

block is FP Growth that means that the data has been read in the process using the FP-Growth Algorithm. The third block is Create Association Rules which means the result of the FP-Growth algorithm is processing to create association rules between a rule that has been resulting from the FP-Growth algorithm.

C. Discussion

According to the data in table 2, based on the results of the K-Means system clustering data processing using the Rapid Miner application, it can be seen that from a total of 16 subject data it is made into three clusters, namely cluster 0 (zero), cluster 1 (one), and cluster 2 (two). Cluster 0 is defined as specialization subject 1, cluster 1 as elective course 2, and cluster 2 as elective course 3. Based on the results of the output, it can be seen that 6 students choose elective subject 1, 5 students who choose elective subject 2, and 5 students who choose elective subjects. elective course 3. In Figure 2, there are details of what subjects are classified as elective 1, specialization 2, and specialization subjects 3. Table 3 shows the centroid value or middle value for each student that was carried out as a data sample. In the FP-Growth method, the number of occurrences of each data is calculated. Following the results of data processing in table 4, the highest data emergence is 3 subjects, namely e-Business, Business Process Restructuring, and Customer Relationship Management. With this value, it can be interpreted that the three subjects are the ones that are taken the most by students. In Figure 3, you can see the values of support, confidence, and Laplace, which are the values from the results of data processing using rapid miner. To produce these values the Rapid Miner application requires a design from the Rapid Miner calculation shown in Figure 4 which consists of 3 blocks, namely read excel data then FP-Growth Algorithm blog, and finally the create association rule block.

IV. Conclusion and Recommendation

Tests were carried out to determine the effect of a rule formed from a group of data based on existing student scores on student decisions in determining the subject of interest to be taken. The input data is in the form of student data with compulsory courses along with their scores with a minimum value of support = 75% and minimum confidence = 80% with a lift ratio ≥ 1 which is used in determining specialization subjects based on their academic abilities. The results of this rule can provide input to students or academic supervisors when compiling student study plan cards. Lecturers and students can analyze the right specialization subject by following the pattern given. More clearly, it can be concluded as follows:

The pattern that is built can produce decision recommendations in determining selected subjects using Clustering and k-means analysis, Association Rule Mining, and FP-Growth algorithms. This decision model can receive dynamic input data according to use policies with a predetermined format. The resulting pattern shows that the elective courses with the theme of business information systems are more followed by students than the other 2 themes. The pattern search process using the FP-growth algorithm is shorter than other association rule algorithms such as the a priori algorithm.

Acknowledgment

Our gratitude goes to the Padang Computer College Foundation (YPTK) which is chaired by the Alm. Mr. H. Herman Nawas and Ibuk Dr. Hj. Zerni Melmusi, S.E., M.M., Ak., C.A. which has provided funding for this research activity.

References

- [1] S. P. Astuti and D. E. Utami, "Development of the Quality of Islamic Higher Education Based on the Desires and Needs of Stakeholders," *JAS-PT J. Anal. Sist. Pendidik. Tinggi*, vol. 2, no. 2, p. 71, 2018, doi: 10.36339/jaspt.v2i2.205.
- [2] Irwansyah, "Implementation of Distance Higher Education System in Private Universities," *JAS-PT J. Anal. Sist. Pendidik. Tinggi*, vol. 2, no. 1, p. 39, 2018, doi: 10.36339/jaspt.v2i1.122.
- [3] I. M. Sadjati and P. R. Pertiwi, "Students' Perceptions of Practicum Implementation in Long Distance Open Higher Education (Case: Open University Faculty of Mathematics and Natural

- Sciences Agribusiness Study Program),” *J. Pendidik. Terbuka dan Jarak Jauh*, vol. 14, no. 1, pp. 45–56, 2013, [Online]. Available: <http://jurnal.ut.ac.id/index.php/JPTJJ/article/view/128/103>.
- [4] T. Darmayanti, M. Y. Setiani, and B. Oetojo, “E-Learning on Distance Education: A Concept That Changes Learning Methods in Higher Education in Indonesia,” *J. Pendidik. Terbuka dan Jarak Jauh*, vol. 8, pp. 99–113, 2007.
- [5] U. Nurhayani, “The Influence of Motivation on the Interest of Accounting Students to Participate in Accounting Professional Education (PPAk) (Empirical Study at Private Universities in Medan),” *J. Mediasi*, vol. 4, no. 1, pp. 59–67, 2012.
- [6] F. Wang, Q. Wang, F. Nie, Z. Li, W. Yu, and F. Ren, “A linear multivariate binary decision tree classifier based on K-means splitting,” vol. 107, 2020, doi: 10.1016/j.patcog.2020.107521.
- [7] Z. Ren, L. Sun, and Q. Zhai, “Int J Appl Earth Obs Geoinformation Improved k-means and spectral matching for hyperspectral mineral mapping,” *Int J Appl Earth Obs Geoinf.*, vol. 91, no. January, p. 102154, 2020, doi: 10.1016/j.jag.2020.102154.
- [8] S. El, W. Fakhret, T. Moulahi, and R. Khan, “Improved node localization using K-means clustering for Wireless Sensor Networks,” *Comput. Sci. Rev.*, vol. 37, p. 100284, 2020, doi: 10.1016/j.cosrev.2020.100284.
- [9] L. Ma and H. Sun, “ScienceDirect Research on Clustering Method Based on Weighted Distance Density and K-Means,” *Procedia Comput. Sci.*, vol. 166, pp. 507–511, 2020, doi: 10.1016/j.procs.2020.02.056.
- [10] S. Bashir, “An Efficient Pattern Growth Approach for Mining Fault Tolerant Frequent Itemsets,” *Expert Syst. Appl.*, p. 113046, 2019, doi: 10.1016/j.eswa.2019.113046.
- [11] P. Kiewit, J. Wang, Z. Cheng, J. Wang, and Z. Cheng, “ScienceDirect FP-Growth based Regular Behaviors Auditing in Electric FP-Growth based Regular Behaviors Auditing in Electric Management Information System Management Information System,” *Procedia Comput. Sci.*, vol. 139, pp. 275–279, 2018, doi: 10.1016/j.procs.2018.10.268.
- [12] L. Bai, J. Liang, and F. Cao, “A multiple k -means clustering ensemble algorithm to find nonlinearly separable clusters,” vol. 61, no. March, pp. 36–47, 2020, doi: 10.1016/j.inffus.2020.03.009.
- [13] W. Cai, J. Zhao, and M. Zhu, “A real time methodology of cluster-system theory-based reliability estimation using k-means clustering,” *Reliab. Eng. Syst. Saf.*, vol. 202, no. March, p. 107045, 2020, doi: 10.1016/j.ress.2020.107045.
- [14] A. P. U. S. Supiyandi, Mochammad Iswan Perangin-angin, Andre Hasudungan Lubis, Ali Ikhwan, Mesran, “Association Rules Analysis on FP-Growth Method in Predicting Sales,” *Int. J. Recent Trends Eng. Res.*, vol. 3, no. 10, pp. 58–65, 2017.
- [15] W. Taktak and Y. Slimani, “MS-FP-Growth: A multi-support Vrsion of FP-Growth Agorithm,” *Int. J. Hybrid Technol.*, vol. 7, no. 3, pp. 155–166, 2014.
- [16] S. Butsianto and N. T. Mayangwulan, “PApplication of Data Mining For Car Sales Prediction Using the K-Means Clustering Method,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 187–201, 2020, doi: 10.32672/jnkti.v3i3.2428.
- [17] K. D. Hartomo, S. Y. Prasetyo, and R. A. Suharjo, “Prediksi Stok dan Pengaturan Tata Letak Barang Menggunakan Kombinasi Algoritma Triple Exponential Smoothing dan FP-Growth,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, p. 869, 2020, doi: 10.25126/jtiik.2020751863.
- [18] Y. A. Auliya, “Improve Hybrid Particle Swarm Optimization and K-Means by Random Injection for Land Clustering of Potato Plants,” *Proc. - 2019 Int. Conf. Comput. Sci. Inf. Technol. Electr. Eng. ICOMITEE 2019*, vol. 4, no. 1, pp. 192–198, 2019, doi: 10.1109/ICOMITEE.2019.8921207.

- [19] P. J. Jones *et al.*, “FilterK : A new outlier detection method for k-means clustering of physical activity,” *J. Biomed. Inform.*, vol. 104, no. October 2019, p. 103397, 2020, doi: 10.1016/j.jbi.2020.103397.
- [20] D. Ranti, A. J. Warburton, K. Hanss, and D. Katz, “K-Means Clustering to Elucidate Vulnerable Subpopulations Among Medicare Patients Undergoing Total Joint Arthroplasty,” *J. Arthroplasty*, 2020, doi: 10.1016/j.arth.2020.06.063.
- [21] B. Wu, D. Zhang, Q. Lan, and J. Zheng, “An efficient frequent patterns mining algorithm based on Apriori algorithm and the FP-tree structure,” *Proc. - 3rd Int. Conf. Converg. Hybrid Inf. Technol. ICCIT 2008*, vol. 1, pp. 1099–1102, 2008, doi: 10.1109/ICCIT.2008.109.
- [22] M. Narvekar and S. Fatma, “An optimized algorithm for association rule mining using FP tree,” *Procedia - Procedia Comput. Sci.*, vol. 45, pp. 101–110, 2015, doi: 10.1016/j.procs.2015.03.097.
- [23] H. Hendri, H. Awal, and Mardison, “Solar-Cell Implementation for Supporting Tourist Facilities and Tourism Promotion Media,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 012058, 2021, doi: 10.1088/1742-6596/1783/1/012058.