THE USE OF HASH-BASED ALGORITHM CALCULATION DEPENDENT ON EXECUTION ASSESSMENT FRAMEWORK TO DECIDE UNDERSTUDY FULFILLMENT WITH EDUCATING STAFF

Dedy Rahman Prehanto, Universitas Negeri Surabaya
Ginanjar Setyo Permadi, Universitas Hasyim Asy'ari
Aries Dwi Indriyanti, Universitas Negeri Surabaya

ABSTRACT

Evaluation of satisfaction is an act of assessing the performance of teaching staff in carrying out their duties or responsibilities and being able to meet predetermined expectations. This study aims to build a system of student satisfaction on the performance of teaching staff and implement a hash based algorithm to help process the satisfaction evaluation assessment. The hash based algorithm generates information for the next iteration in the current iteration, so that it can efficiently generate large itemset and can be a solution in determining the frequent itemset from candidate itemset optimally. In this study, the data is processed using a hash-based algorithm in order to obtain results in the form of frequent itemsets that are useful in forming association rules. From the results of the formation of these association rules, it can be analyzed the assessment pattern of the teaching staff, in order to be a morale booster, or improve the quality of the teaching staff. The results showed that the index of the level of student satisfaction on the performance of teaching staff in the information system design course ranged from 54% to 100%. This means that students are satisfied at the moderate to very good level.

Keywords: Association Rules; Hash Based; Student Satisfaction; Teacher Performance.

INTRODUCTION

At present the community's need for formally held education is increasing, especially in universities, this can make universities as producers of quality human resources. Improving the quality of a university is influenced by the quality of the university. Factors that influence the quality of education are teaching staff who play a very important role in producing skilled and quality graduates (Kistofer et al., 2019).

In this study, the instrument of student satisfaction on the performance of teaching staff was taken from the Quality Assurance Institute of Hasyim As'ari University using four criteria, namely, learning preparation, learning process, learning evaluation and scientific attitude. Satisfaction is a positive attitude or perception of students towards products or services that have met their expectations, students will feel dissatisfied if their expectations have not been achieved. A service has the potential to meet or not the expectations of students. Assessment of the level of student satisfaction is one aspect of the assessment of the quality of educational services which has an important meaning for the continuity of the education system. The results of these assessments can be used to guide a better education system (Ghazanfari et al., 2020).

The results of data processing in a simple way do not get effective results because of the large volume of data processed. Utilization of information and knowledge contained in the amount of data is called data mining (Mashuri et al., 2019). Data mining is a series of processes that use statistical, mathematical and artificial intelligence techniques to convert useless data into data that contains valuable and useful information from large databases (Dzakiyullah et al., 2018; Aries Dwi Indriyanti et al., 2019a; Wardhana et al., 2019). Data mining is not a new field, data mining inherits many aspects and techniques from existing fields of science (Prehanto et al., 2020).

One of the data mining algorithms that is often used in association rules is a hash based algorithm. The hash based algorithm produces information for the next iteration in the current iteration, so that it can efficiently generate large itemset and can be a solution in determining the frequent itemset from candidate itemset optimally (Prehanto et al., 2019). The hash based algorithm uses hashing techniques to filter out itemsets that are not important for the next itemset generation (Permadi et al., 2019).

MATERIALS AND METHODS

Data Mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases (Zahrotun et al., 2018). Data mining is divided into several based on the tasks that can be performed, namely (Si et al., 2019):

1. Description: The analysis simply wants to try to find a way to describe the patterns and trends contained in the data. Patterns and tendencies often provide possible explanations for a pattern or trend.
2. Estimate: In the estimation process, it tends to be numerical rather than nominal. The model is built using a complete record that provides the value of the target variable as the predicted value. Examples of these algorithms are Linear Regression and Neural Network.
3. Prediction: In the prediction, the data used is time series data and the final result value is used for some time to come.
4. Classification: In the classification of the target variable in the form of categorical data. An example of classification is people's income is classified into three groups, namely high, medium and low income. The classification algorithms are Naïve Bayes, K-Nearst Neighbor, and C4.5.
5. Clustering: Grouping of data or the formation of data in the same type. Clustering divides data into relatively equal groups. Examples of clustering algorithms are Fuzzy C-Means and Kmeans.
6. Association: Associations are used to find items that appear at the same time and look for relationships between one or two more in a data set. Examples of association algorithms are FP-Growth, Apriori and Hash Based (Hossain et al., 2019).

In this study the authors used association groups. Association is the process of processing data to find a combination of items that appear at one time (Indriyanti et al., 2019b).

Association Rule

Association Rule is a process to find patterns of associative rules that meet the minimum requirements for support and the minimum requirements for confidence in the database. Association rule is a procedure to find the relationship between items in a specified dataset. Association rule includes two stages(Nomura et al., 2020). The stages are:

1. Look for the most frequent combination of an item set that has support above the minimum support. This is called the frequent itemset.
2. Defines association rules from frequent itemset that have been created based on the minimum support and minimum confidence rules.

Generally, there are two measures used in determining an association rule, namely support and confidence. These two measures will later be useful in determining the attractive association rule, which is to be compared with a predetermined limit. The limitation consists of minsup and minconf (Prehanto et al., 2019).

Minimum Support

Minimum support is a measure or value that must be met as a limit on the frequency of occurrence (support count) of the entire dominance value of an item or itemset (support) in the entire transaction.

$$\text{Support A} = \frac{\text{Amount Transaction of containing A}}{\text{Total Transaction}} \text{ x } 100\% \qquad (2.1)$$

To find the support value of 2 items is as follows:

$$\text{Support (A,B)} = \frac{\text{Amount Transaction of containing A and B}}{\text{Total Transaction}} \text{ x } 100\% \qquad (2.2)$$

Minimum Confidence

The minimum confidence value is a parameter that defines the minimum level of a relationship value between items (confidence) that must be met in order to find quality rules. The confidence value is obtained by the following formula:

$$\text{Confidence} = P\,(A|B) \frac{\text{Amount Transaction of containing A and B}}{\text{Amount Transaction of containing A}} \text{ x } 100\% \qquad (2.3)$$

Hash Based Algorithm

Hash based algorithm is the development of a priori algorithm which is used to determine rule association. Hash based algorithm can be a solution to overcome the problem of determining the frequent k-itemset from a large number of k-itemset candidates. By using hashing technique, the scanned k-itemset candidates will be entered into the bucket in the hash table. From the bucket, it will be used to search for frequent (k+1)-itemsets so that the database scan is only done once in the first iteration (Siregar et al., 2021). This algorithm generates information for the next iteration in the current iteration. The technique used in the hash based algorithm is a hashing technique to filter out itemsets that are not important for the next itemset generation. When the support count for candidate itemsets is calculated by browsing the database, this algorithm collects information about the (k+1)-itemsets by means of all possible (k+1)-itemsets being hashed into a hash table using a hash function. The bouquet in each hash table contains the number of times the itemset has been hashed into the bouquet (Imdadi & Rizvi, 2011). Based on the hash table, then a bit vector will be built, where the bit vector is worth 1 if the number in the bouquet is greater than or equal to the minimum support. In the candidate generation section, after calculating $Ck = Lk1 * Lk - 1$, each k-itemset is checked whether the itemset is hashed into a bouquet that has a bit vector equal to 1. Otherwise, the vector will be omitted or not used. The use of a hash table reduces the number of candidate k-itemsets, thereby reducing the computational value of generating

the itemset in each iteration (Naresh & Suguna, 2019). The process of hashing the candidate 1-itemset using the formula (Naresh & Suguna, 2019):

$$h(x) = (\text{order of item } x) \bmod n \qquad (2.4)$$

Where, h = bucket address in hash table
n = multiple addresses (n=2.m+1, m=total number of items)
Process hashing against candidates 2-itemset:

$$h(k) = ((\text{order of } x) * 10 + \text{order of } y) \bmod n \quad (2.5)$$

In the calculation of the 2-itemset formula, a collision will be found (there is more than 1 itemset that has the same hash address). When a collision occurs, it is immediately checked to find the address of an empty bucket. After checking, if an indication is found that the hash table is half filled, it is necessary to rehash it with 2 times the number of addresses before. To solve this problem, a different formula is used, namely(Ali et al., 2019):

$$h(k) = ((\text{order of } x) * 10 + \text{order of } y) \bmod j \quad (2.6)$$

Where,
j : many addresses after adding (j=2*m+1), m= the number of addresses in the hash table before adding)
Process hashing against candidates 3-itemset:

$$h(k) = ((\text{order of } x) * 100 + (\text{order of } y) * 10 + \text{order of } z) \bmod j \qquad (2.7)$$

RESULTS AND DISSCUSION

The instrument used in this study is intended to produce accurate data by using a Likert scale. The Likert scale is used to measure an attitude, opinion and perception of a person or group of people about a social phenomenon(Permadi et al., 2018). Selection of the right instrument is needed to obtain objective data and information. The instrument used in this study is an instrument for assessing the performance of teaching staff at Hasyim Asy'ari University (Table 1).

No	Table 1 TEACHER PERFORMANCE ASSESSMENT INSTRUMENT
	Kind of service
	A. Learning Preparation
1	Submitting Semester Lecture Plans (RPS)
2	Submit a study agreement/contract
3	Number of attendance of substitute teachers/lectures
	B. learning process
4	Delivering lecture material clearly
5	Provide opportunities for students to ask questions / express opinions
6	Teachers' answers to students' questions related to lecture material materi
7	Lecture time according to Semester Credit Unit weight Kredit (SKS)
8	Keeping the class schedule according to the schedule determined majors / study programs
9	Optimal use of facilities and infrastructure
10	Return of Assignments and Mid-Semester Exams (UTS)
11	Suitability of lecture/practice materials, assignments, UTS and UAS
12	Delivering information on various source books/ textbooks/ handouts
13	Conducting tutoring / structured
14	Has a variety of up-to-date references
	C. Learning Evaluation

15	Objectivity in assessment: assignments, tests, and lecture participation
16	Punctuality in submitting grades
	D. Scientific Attitude
17	The attitude of openness of scientists and academics
18	Dress politely, neatly and appropriately
19	Easy to find for related consultation study
20	Ability to create a conducive atmosphere/ motivation suasana

The flowchart design of the student satisfaction system on the performance of the teaching staff can be seen in the Figure 1.
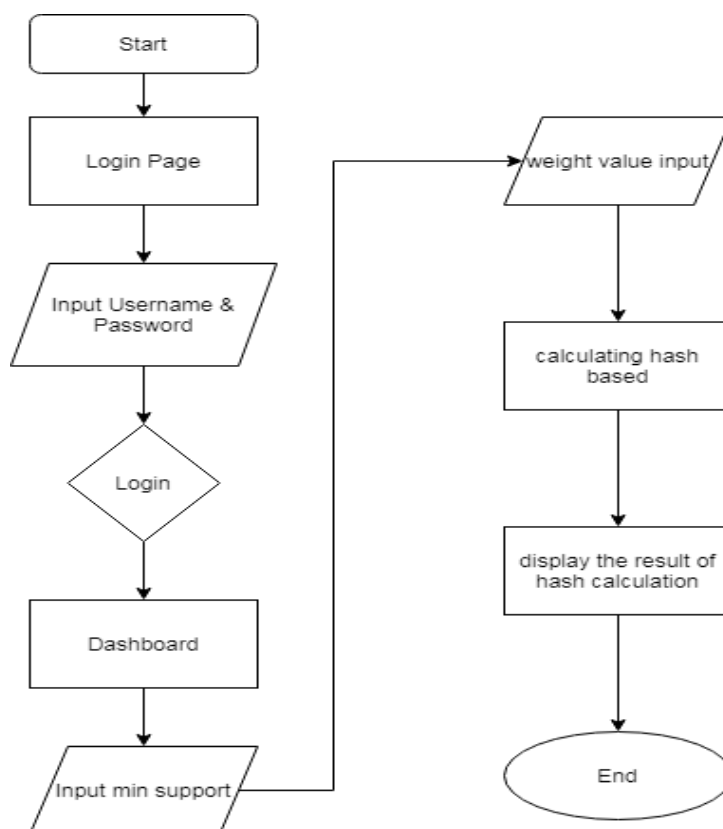

FIGURE 1

FLOWCHART DESIGN OF THE STUDENT SATISFACTION SYSTEM

Pre-Processing Data

This is the stage where data initialization is carried out before the data is processed using a hash based algorithm. The data initialization process is carried out by converting the initial data which is still in the form of assessment questionnaire data into data that is ready to be processed by a hash-based algorithm, the data that is ready to be processed can be called transaction data. The following is the initial data before the preprocessing stage is mapped in the Table 2.

The data used is assessment data on aspect 1 from one of the lecturers at Hasyim Asy'ari University, with a minimum support of 50% and a minimum confidence of 50%. The meaning of the numbers in Table 2 is the questionnaire rating scale of the teaching staff.

Table 2 FIRST DATA										
Question Answer	11	12	13	14	15	16	17	18	19	10
1	5	5	5	4	4	5	4	5	5	5
2	5	4	3	5	4	3	3	5	5	4
3	3	5	4	4	4	4	4	5	3	4
4	4	3	3	4	4	5	5	4	5	4
5	5	4	3	4	5	4	4	5	5	4
6	5	4	4	4	5	4	3	4	4	4
7	3	4	3	4	5	5	5	4	4	3
8	4	4	4	5	4	4	4	3	4	4
9	3	3	3	4	4	3	4	5	2	4
10	5	4	4	4	5	4	3	4	4	4
11	3	4	3	4	5	5	5	4	4	3
12	4	4	4	5	4	4	4	3	4	4
13	3	3	3	4	4	3	4	5	2	4
14	4	3	3	4	3	4	4	3	4	3
15	3	3	3	4	4	4	3	4	2	4
16	4	3	3	4	4	4	4	4	3	4
17	3	3	3	4	3	4	3	3	3	4
18	4	4	3	4	3	3	4	4	4	3
19	3	3	3	4	5	5	5	4	4	4
Question Answer	11	12	13	14	15	16	17	18	19	10
1	4	4	3	4	4	4	5	4	3	3
2	4	3	4	5	4	3	3	5	4	4
3	5	4	4	4	3	4	3	3	4	4
4	5	5	4	3	5	5	4	4	4	3
5	4	4	3	5	5	4	4	3	4	4
6	3	4	4	5	4	4	3	4	3	4
7	4	5	4	4	5	4	4	5	4	5
8	5	3	2	3	2	4	3	4	2	3
9	4	4	2	4	4	4	4	4	3	4
10	3	4	4	5	4	4	3	4	3	4
11	4	5	4	4	5	4	4	5	4	5
12	5	3	2	3	2	4	3	4	2	3
13	4	4	2	4	4	4	4	4	3	4
14	4	3	3	3	4	3	3	4	4	3
15	3	3	4	4	4	4	4	4	3	3
16	4	4	4	3	4	5	3	3	4	3
17	4	3	3	3	3	3	4	3	4	4
18	3	3	3	3	4	4	3	3	4	3
19	4	5	5	5	4	3	5	3	3	4

Very good=5; Good=4; Enough=3; Less=2; Very Less=1

After preprocessing, it will get ready-to-process data in the form of transaction data that is ready to be processed by a hash-based algorithm, which is tabled in Table 3.

Calculation Hash-based Algorithm

The ready data that has been formed in the preprocessing stage is then processed using a hash based algorithm in order to obtain results in the form of frequent itemsets that are useful in forming association rules. There are 2 stages used in the hash-based algorithm, namely the formation of frequent items and the formation of association rules, here are the steps:

Formation of Frequent Item Set Aspect 1

The formation of frequent itemset is the main stage in the association rules, because at this stage the formation of itemsets is carried out which includes mapping and itemset combinations, itemset hashing addressing, itemset pruning, which from the whole process is influenced by minimum support. In the formation of frequent itemset, transaction data obtained from the preprocessing stage is required which is mapped in Table 3.

Table 3
TRANSACTION DATA

Transaction ID	Item	Transaction ID	Item
T1	SB,SB,SB	T13	B,K,C
T2	SB,B,C	T14	C,C,B
T3	C,SB,B	T15	C,C,B
T4	B,C,C	T16	C,C,C
T5	SB,B,C	T17	B,C,C
T6	SB,B,B	T18	C,C,C
T7	C,B,C	T19	B,B,C
T8	B,B,B	T20	C,C,C
T9	C,C,C	T21	B,B,SB
T10	B,C,C	T22	K,C,C
T11	SB,B,B,	T23	C,C,C
T12	SB,C,B	T24	C,C,B

Itemset Mapping (1-itemset)

From 24 transaction data, 4 itemsets were obtained which are tabled in Table 4.

Table 4
ITERATION ITEMSET MAPPING 1

No	Itemset
1	B
2	C
3	K
4	SB

Hashing Addressing (1-itemset)

To perform hasing, each itemset is ordered by order of items in the data. The order of items in the data is used as the first address sequence before further processing. For more details can be seen in Table 5.

Table 5
ORDER OF ITEM

Order of item	Itemset
1	B
2	C
3	K
4	SB

After sorting the 1-itemset itemset, then the hashing process is carried out by entering each itemset into the bucket in the hash table using equation 2.4, namely $h(x) = (order\ of\ item\ x)\ mod\ n$ dengan n= 9.

Table 6
PROSES HASHING 1-ITEMSET

Itemset	Process	Address
B	h(B) = (1) mod 9	1
C	h(B) = (2) mod 9	2
K	h(C) = (3) mod 9	3
SB	h(SB) = (4) mod 9	4

Trimming Item Set (1-itemset)

Itemset pruning is used to filter itemsets that match the specified minimum support, which is 50% of the transaction data or equal to 12 pieces, if less than 12 items then the itemsets will be removed and not used in the next iteration. The following itemsets will be carried out in the 1-itemset pruning process, which can be seen in the Table 6.

Table 7
TABLE JOIN TRIMMING

Item set	Count
B	17
C	20
K	2
SB	7

After trimming, the data from the trimming will be generated as shown in Table 7.

Table 8
PRUNING RESULTS TABLE

itemset	count
B	17
C	20
K	2
SB	7

With the pruning results obtained as shown in Table 8, there are several itemsets that meet the minimum support, so it can be continued to the next stage, by generating the itemset again, namely 2-itemset.

Itemset Mapping (2-Itemset)

The 4 itemsets in the first iteration (1-itemset) 2 itemsets were obtained that passed the second iteration (2-itemset), and a combination process was carried out which resulted in 1 itemset (Table 9).

Table 9
ITEMSET MAPPING 2-ITEMSET

No	Itemset
1	B,C

Addressing Hashing (2-itemset)

Then the hashing process is carried out on the 2-itemset itemset by entering each itemset into the bucket in the hash table. The hashing process on itemsets that have more than

one itemset (1-itemset) will use equation 2.5, namely h(k)=((order of x) * 10 + order of y)mod n with n=3. $h(B,C) = ((1) * 10 + 2 \, mod \, 3 = 0$ After addressing, the address of the itemset 2 itemset is obtained which is tabled in Table 10.

Table 10
ADDRESS 2-ITEMSET ITERASI 1

Itemset	Alamat
B,SB	0

After that, the frequent itemset collection process is carried out from the hash-based calculation process that has been carried out. Can be seen in the Table 11.

Table 11
FREQUENT ITEMSETS

Item	Count	Support
B	17	(17/24) x 100 = 70%
C	20	(20/24) x 100 = 83%
B,C	14	(14/24) x 100 = 58%

Establishment of Association Rules Aspects 2

The formation of Association rules is used to determine the relationship between the itemset, where the itemset in the form of a scale will show the relationship between which scales are said to be good from the assessment of the teaching staff concerned. From the results of the formation of these association rules, it can be analyzed the assessment pattern of the teaching staff, in order to be a morale booster, or improve the quality of the teaching staff. The results of the formation of association rules with a minimum 50% confidence can be seen in Table 12.

Table 12
ASOCIATION RULES

Item	Support (XY)	Support (X)	Confidence
B ≥ C	14	17	0,82
C ≥ B	14	20	0,70

Repeat these steps on aspects 2,3 and 4.

Displays all the results of association rules from all aspects.

After all aspects are calculated using a hash based, then the next step is to collect all the association rules from the results of each existing aspect. The following is the result of the overall association rule (Tables 13-15).

Table 13
ASPECT ASSOCIATION RULES 1

Item	Support (XY)	Support (X)	Confidence
B ≥ C	14	17	0,82
C ≥ B	14	20	0,7

Table 14
ASPECT ASSOCIATION RULES 2

Item	Support (XY)	Support (X)	Confidence
B ≥ C	23	24	0,95
B ≥ SB	14	24	0,58
C ≥ B	23	23	1
C ≥ SB	13	23	0,56
SB ≥ B	14	14	1
SB ≥ C	13	14	0,92
B ≥ C,SB	13	24	0,54
C ≥ B,SB	13	23	0,56
SB ≥ B,C	13	14	0,92
B,C ≥ SB	13	23	0,56
B,SB ≥ C	13	14	0,92
C,SB ≥ B	13	13	1

Table 15
ASPECT ASSOCIATION RULES 4

Item	Support (XY)	Support (X)	Confidence
B ≥ C	18	21	0,85
C ≥ B	18	21	0,85

As a result, 16 process association rules were formed based on predetermined parameters, namely a minimum support of 50% and a minimum of 50% confidence. In one of the rules formed, for example the rule in aspect 2: B ≥ C with a 95% confidence value means that 95% of students who choose good also choose enough.

CONCLUSION

Based on the research that has been done, the conclusion of this study is that the index of the level of student satisfaction on the performance of teaching staff in information system design courses ranges from 54% to 100%. This means that students are satisfied at the level of enough to very good.

REFERENCES

Ali, Y., Farooq, A., Alam, T.M., Farooq, M.S., Awan, M.J., & Baig, T.I. (2019). Detection of Schistosomiasis Factors Using Association Rule Mining. *IEEE Access, 7*, 186108-186114.

Dzakiyullah, N.R., Saleh, C., Rina, F., & Fitra, A.R. (2018). Estimation of Carbon Dioxide Emission Using Adaptive Neuro-Fuzzy Inference System. *Journal of Engineering and Applied Sciences, 13*(6SI), 5196-5202.

Ghazanfari, B., Afghah, F., & Taylor, M.E. (2020). Sequential Association Rule Mining for Autonomously Extracting Hierarchical Task Structures in Reinforcement Learning. *IEEE Access, 8, 11782-11799.*

Hossain, M., Sattar, A.H.M.S., & Paul, M.K. (2019). Market basket analysis using apriori and FP growth algorithm. *2019 22nd International Conference on Computer and Information Technology, ICCIT 2019, 1-6.*

Imdadi, N., & Rizvi, S.A.M. (2011). Using Hash based Bucket Algorithm to Select Online Ontologies for Ontology Engineering through Reuse. *International Journal of Computer Applications, 28*, 21-25.

Indriyanti, A.D., Prehanto, D.R., Permadi, G.S., Mashuri, C., & Vitadiar, T.Z. (2019a). Using Fuzzy Time Series (FTS) and Linear Programming for Production Planning and Planting Pattern Scheduling Red Onion.

The 4th International Conference on Energy, Environment, Epidemiology and Information System (ICENIS 2019).

Indriyanti, A.D., Prehanto, D.R., Prismana, I.G.L.E.P., Soeryanto, Sujatmiko, B., & Fikandda, J. (2019b). Simple Additive Weighting algorithm to aid administrator decision making of the underprivileged scholarship. *Journal of Physics: Conference Series*, 1402.

Kistofer, T., Permadi, G.S., & Vitadiar, T.Z. (2019). Development of Digital System Learning Media Using Digital Learning System. *Proceedings of the 1st Vocational Education International Conference (VEIC 2019)*.

Mashuri, C., Mujianto, A.H., Sucipto, H., Arsam, R.Y., & Permadi, G.S. (2019). Production Time Optimization using Campbell Dudek Smith (CDS) Algorithm for Production Scheduling. *The 4th International Conference on Energy, Environment, Epidemiology and Information System (ICENIS 2019)*.

Naresh, P., & Suguna, R. (2019). Association rule mining algorithms on large and small datasets: A comparative study. *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*, 587-592.

Nomura, K., Shiraishi, Y., Mohri, M., & Morii, M. (2020). Secure Association Rule Mining on Vertically Partitioned Data Using Private-Set Intersection. *IEEE Access, 8*, 144458-144467.

Permadi, G.S., Adi, K., & Gernowo, R. (2018). Application Mail Tracking Using RSA Algorithm As Security Data and HOT-Fit a Model for Evaluation System. *The 2nd International Conference on Energy, Environmental and Information System (ICENIS 2017)*.

Permadi, G.S., Vitadiar, T.Z., Kistofer, T., & Mujianto, A.H. (2019). The Decision Making Trial and Evaluation Laboratory (Dematel) and Analytic Network Process (ANP) for Learning Material Evaluation System. *The 4th International Conference on Energy, Environment, Epidemiology and Information System (ICENIS 2019)*.

Prehanto, D.R., Indriyanti, A.D., Nuryana, K.D., Soeryanto, S., & Mubarok, A.S. (2019). Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token. *Journal of Physics: Conference Series*, 1402.

Prehanto, D.R., Indriyanti, A.D., Permadi, G.S., Vitadiar, T.Z., & Jayanti, F.D. (2020). Library book modeling data using the association rule method with apriori algorithm in determining book placement and analysis of book loans. *International Journal of Advanced Science and Technology*, *29*(5), 1244-1250.

Si, H., Zhou, J., Chen, Z., Wan, J., Xiong, N.N., Zhang, W., & Vasilakos, A.V. (2019). Association Rules Mining among Interests and Applications for Users on Social Networks. *IEEE Access*, *7*, 116014-116026.

Siregar, A.H., Lydia, M.S., & Wage, S. (2021). Association Rule Analysis using CT-Pro and Hash-based Algorithm in Violence Case of Children. *In Proceedings of the International Conference on Culture Heritage, Education, Sustainable Tourism, and Innovation Technologies - CESIT, 565-573*.

Wardhana, M.H., Samad, A., Basari, H., Syukor, A., Jaya, M., & Afandi, D. (2019). A Hybrid Model using Artificial Neural Network and Genetic Algorithm for Degree of Injury Determination. *International Journal of Innovative Technology and Exploring Engineering, 9*(2), 1357-1365.

Zahrotun, L., Soyusiawaty, D., & Pattihua, R.S. (2018). The implementation of data mining for association patterns determination using temporal association methods in medicine data. *2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018*.