

Normalized Data Technique Performance for Covid-19 Social Assistance Decision Making

* case: student's internet data social assistance during learning from home due covid19

1st Edy Budiman
Dept. of Informatics
Universitas Mulawarman
Samarinda, Indonesia
edy.budiman@fkti.unmul.ac.id

2nd Joan Angelina Widians
Dept. of Informatics
Universitas Mulawarman
Samarinda, Indonesia
angel.unmul@gmail.com

3rd Masna Wati
Dept. of Informatics
Universitas Mulawarman
Samarinda, Indonesia
masnawati.ssi@gmail.com

4th Novianti Puspitasari
dept. of informatics
Universitas Mulawarman
Samarinda, Indonesia
novia.ftik.unmul@gmail.com

Abstract—The student internet data assistance program is an effort by educational institutions to support online learning from home during the Covid-19 pandemic. A series of tests are applied to determine the optimization of decision making on the social assistance program performance. This study aims to evaluate the performance of students' internet data assistance programs using a confusion matrix approach, in particular on the performance of simple, linear and vector normalized data analysis techniques. The representation normalized techniques performance for simple data using SAW, linear data is VIKOR and vector using the MOORA method. The study results found that there were differences in performance in the process of selecting preferences for ranking potential social assistance recipients, as well as a differential in the confusion matrix performance values on the accuracy, precision, recall and error rate values on each method.

Keywords—confusion matrix, normalized data, social assistance, decision making.

Keywords—component, formatting, style, styling, insert (key words)

I. INTRODUCTION

The Covid-19 pandemic impact's, Indonesian government issued a Large-Scale Social Restriction (PSBB) policy to reduce the spread of Covid-19[1],[2]. Related to this, the Ministry of Education and Culture issued a circular on the Implementation of Education Policy in Emergency during COVID-19 Spread[3]. This circular, among other things, contains policies regarding social distancing through working and learning from home(LFH).

The policy of keeping people stay at home automatically makes internet service crucial and high service. Many workers use the internet to work, include in the education sector, schools or education institutions that use an online communication media as an alternative to conventional learning. However, the use of the internet is a problem for most students, the reason is that during the LFH period, the students parents were less able to finance or increase the budget for purchasing internet data packages for their children.

One of the efforts of some local governments and educational institutions in helping students is the policy of the internet data assistance program to support online learning from home, this program is distributed every month with the

same data package value. This effort is certainly very helpful for students. However, the author's subjective view is that the decision making in determining internet data packages should be through a needs analysis approach, based on indicators of the amount of internet data usage in online learning and the ability of the students' economic costs.

For this reason, the problem statement in this, we study propose of the implementation methods in data management for Covid-19 social assistance program decision-making through a multi-criteria analysis method approach. In particularly normalized data techniques.

This study aims to evaluate the performance of students' internet data assistance programs using a confusion matrix approach, in particular on the performance of simple, linear and vector the normalized data techniques. The representation of normalized data techniques performance for simple normalized technique is SAW method, linear normalized technique using VIKOR method and vector normalized technique using the MOORA method. The confusion matrix performance model testing is evaluated against the accuracy, precision, recall and error rate values of each normalized techniques.

Normalized technique affects the results in multiple-criteria decision analysis (MCDA)[4], data normalized according to Vafaei et al [5]is essential for decision-making methods because data has to be numerical and comparable to be aggregated into a single score per alternative[5]. Various models approaches are applied in the data normalized technique process as previously reviewed in [6], selection of normalization technique for weighted average multi-criteria decision-making, an integrated multi criteria decision making for a destitute problem[7], and others. Therefore, it is a challenge to select a suitable normalized technique and appropriate handling in case of internet data assistance programs.

The research contribution is directed towards a proposed approach decision making analysis related to normalized data techniques that influence preference assessments, and in general on the analysis of decision-making data for social assistance cases.

II. RELATED WORK

A. Normalized Data Techniques

Data normalization techniques studied in MADM, and is the most important part in determining preferences. This research study has been discussed and applied in several studies, such as:

Nazanin Vafaei et al, on the assessment approach to evaluate normalization techniques using the TOPSIS method[5], and Weighted Average (WA) or SAW (Simple Additive Weighting)[6]. A. Jahan and K. Edwards an state-of-the-art survey on the influence of normalization techniques in ranking for improving the materials selection process in engineering design[8], Investigating the effect of normalization norms in flexible manufacturing sytem selection using multi-criteria decision-making methods[9] by P. Chatterjee and S. Chakraborty, etc.

The implementation the normalized data technique in this research case, we used 3 different approaches, namely simple normalization (SAW), linear (VIKOR) and vector (MOORA), besides that it focused on evaluating the results of the assessment, also testing the performance of the three techniques using the confusion matrix. to the value of accuracy, precision, recall and error rate.

B. Internet Data Assistance Program

An example of an internet data assistance program in this study takes the case of a higher education institution in East Kalimantan, In Mulawarman University. Specifically in the Undergraduate Informatics department. The number of informatics students is very large, and they come from a lower average economic capacity. As a result of the Covid-19 pandemic, it has certainly caused additional living costs for their parents. So that the existence of internet data assistance programs by the Mulawarman University will reduce the burden and support student in online learning.

TABLE I. DATA INTERNET ASSISTANCE PROGRAM CRITERIA

Code	Criteria		
	Criteria	Straight Rank	Attribute
C1	Internet data usage	1	Max
C2	Credit course	2	Max
C3	Economic cost	3	Min

Initial data collection is done through field observations, this activity is in the form of online questionnaires and internet data measurements during the learning process which the author has discussed in the study in [10] Mobile data usage on online learning during Covid-19 pandemic in higher education. The results of this activity determine three (3) main criteria along with the level of importance of weight shown in “Table I”.

III. DATA ANALISYS METHODOLOGY

This section briefly presents the methods used in data analysis of internet data assistance programs. General, the data analysis process design methodology is described in “Fig 1”.

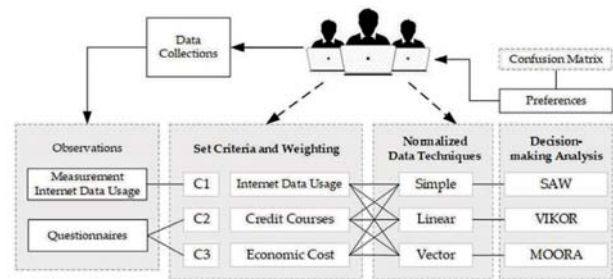


Fig. 1. Design data analysis for normalized techniques

A. Data Collection Methods

The data collection method uses field observation, which involves a series of data measurement activities on the internet and online questionnaire techniques. This activity involved 300 samples of undergraduate students in the Informatics department of Mulawarman University. The collected data is used as the basic parameter in determining social assistance.

B. Criteria Data and Weights

Criteria data are obtained from the results of observation activities, which are seen in “Table I”, criteria ie.; Internet data usage (C1), credit courses (C2) and economic costs (C3). In determining the weighting method, we use the ranking weighting technique Rank Sum approach referring to equation “(1)” from M. Danielson and L. Ekenberg[11],[12] as below;

$$w_i^{RS} = \frac{N+1-i}{\sum_{j=1}^N (N+1-j)} \quad (1)$$

Denote the ranking number i among N items to rank, a larger weight is assigned to lower ranking numbers[11].

For the assignment of importance weight, criteria C1 (internet data usage) and C2 (credit course) are assigned the weight for benefit (max) with Straight rank (r_j) 1 and 2, and C3 (economic cost) has the value attribute Cost(min).

C. Normalized Data Techniques

The normalized data technique used follows the equation “(2)” from D Pavlicic[4], for simple normalization, and the linear equation “(3)” from Nazanin Vafaei et al[6], and for vector refers to the A. Jahan and K.L Edwards[8] in equation “(4)”, which is presented as follows:

- Simple:

$$S_{max} = \frac{x_{ij}}{x_j^*}; S_{min} = \frac{x_j^-}{x_{ij}} \quad (2)$$

$$x_j^* = \max_i x_{ij}, x_j^- = \min_i x_{ij}$$

- Linear:

$$L_{max} = \frac{a_{ij} - a_j^{min}}{a_j^{max} - a_j^{min}}; L_{min} = \frac{a_j^{max} - a_j^{min}}{a_j^{max} - a_j^{min}} \quad (3)$$

- Vector:

$$V_{max} = \frac{r_{ij}}{\sqrt{\sum_{i=1}^m r_{ij}^2}}; V_{min} = 1 - \frac{r_{ij}}{\sqrt{\sum_{i=1}^m r_{ij}^2}} \quad (4)$$

D. Decision-making Analysis

Simple, linear and vector normalized data analysis techniques are illustrated in several analysis methods in

preference decisions. In this study, a simple technical illustration uses the equation "(5)" SAW from [13], Linear with Vikor advancing the equation "(6)" from [14], and for vectors using MOORA refers to equation "(7)" from [15], [16]. The equation used is shown as follows:

- SAW[13]:

$$V_i = \sum_{j=1}^n w_j r_{ij} \tag{5}$$

- VIKOR[14]:

$$Q_1 = v \left[\frac{s_1 - s^-}{s_+ - s^-} \right] + (1 - v) \left[\frac{R_1 - RS^-}{s_+ - RS^-} \right] \tag{6}$$

- MOORA[15]:

$$y_i \sum_{j=1}^g w_j x_{ij}^* - \sum_{j=g+1}^n w_j x_{ij}^* \tag{7}$$

The preference results of the three methods are ordered from the highest to the lowest values, and this is the performance value of the simple, linear and vector normalized data techniques.

E. Confusion Matrix

The confusion matrix is used to measure the performance value of 3 normalized data methods, the equation of this method refers to [17][18],[19], ie.:

- Accuracy : $\frac{TP+TN}{TP+TN+FP+FN}$ (8)

- Precision : $\frac{TP}{TP+FP}$ (9)

- Recal : $\frac{TP}{TP+FN}$ (10)

- Error rate : $\frac{FP+FN}{TP+TN+FP+FN}$ (11)

The performance of the SAW, VIKOR and MOORA methods as an illustration of the performance of simple, linear and vector normalized techniques is compared with the actual data that is the priority target in the internet data assistance program. The actual data we divided into 2 groups, with the number of targets (alternatives) for first priority (I) as many as 40 students, and for priority target II as many as 60 students, out of 300 total students.

IV. RESULT AND DISCUSSION

We present 2 parts in this section, the first part presents the results of data analysis briefly from each process of data statistics, calculating the data normalization method, preference and confusing matrix testing, the second part is discussion.

A. Data Statistics

The statistical data per criteria presented, for the criterion statistical data (C1) in "Table II" is the measurement result data of students' internet data usage when learning online using video communication media.

TABLE II. STATISTICS DATA

Metric	Internet data usage	Credit courses	Economic costs
N	300	300	300
Mean	735.00	20.32	2160816.67
Median	719.99	21.00	1975000
Mode	490.56	24.00	1650000

Std. Dev.	144.017	3.14193	655610
Minimum	490.56	14.00	1000000
Maximum	1130.17	24.00	4000000

The average data usage of students is 735 Mb per course, for the C2 criterion in "Tabel II", the average number of credits is 20.32 courses. Meanwhile, statistics data for the C3 criterion, the average economic cost capability is IDR 2,160,816.67. This shows that the economic cost ability of students is below.

B. Results: Data Normalized Analysis Technique

1) Data normalized for simple technique: based on the results of calculations using equation "(2)", the value of each technique is obtained as shown in "Fig. 2".

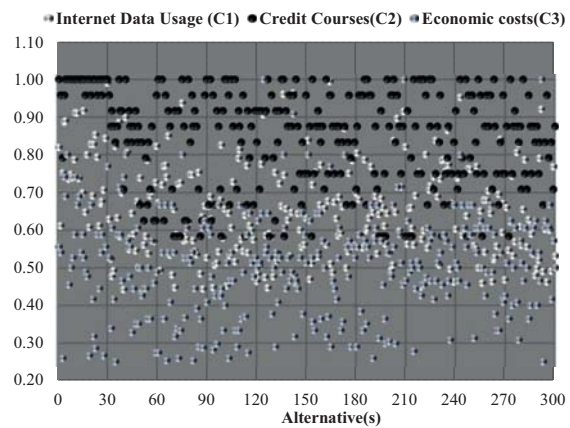


Fig. 2. Scatter-chart distributions of data normalized simple

The simple technique normalized data distribution is seen in the scatter chart "Fig. 2", shows that the value in the range of scatter areas [0.2 to 1], with the centered spread area of C3 criteria [0.4 to 0.6], C2 criteria in areas [0.6 to 0.8] and for C1 in the spread area of the range [0.8 to 1].

2) Data normalized for linear technique: based on the results of calculations using equation "(3)", the value of each technique is obtained as shown in "Fig. 3".

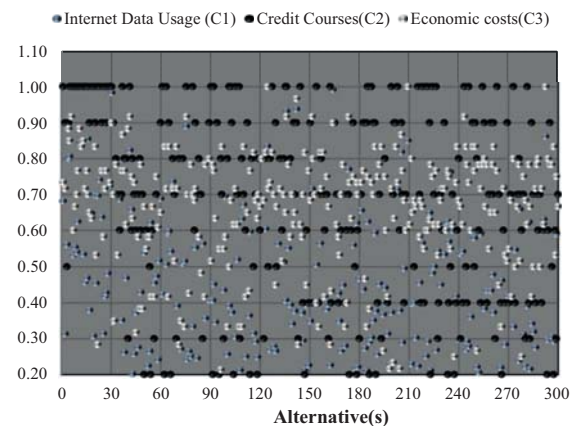


Fig. 3. Scatter-chart distributions of data normalized linear

The linear technique normalized data distribution is seen in the scatter chart "Fig. 3", shows that the value in the range of scatter areas [0.2 to 1], with the centered spread area of C3 criteria [0.2 to 0.9], C2 criteria in areas [0.2 to 1.0] and for C1 in the spread area of the range [0.2 to 0.8]

3) *Data normalized for vector technique*: based on the results of calculations using equation "(4)", the value of each technique is obtained as shown in "Fig. 4".

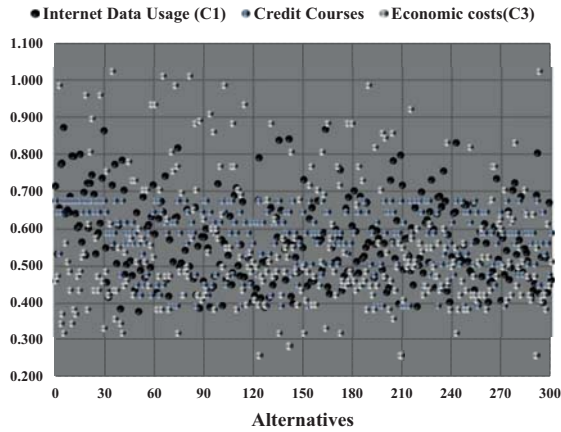


Fig. 4. Scatter-chart distributions of data normalized vector

The vector technique normalized data distribution is seen in the scatter chart "Fig. 4", shows that the value in the range of scatter areas min-max [0.25 to 1.02], with the spread area of C3 criteria [0.25 to 1.02], C2 criteria in areas [0.4 to 0.67] and for C1 in the spread area of the range [0.37 to 0.87]

C. *Data Analysis Decision-making for Preference*

Preference value calculation for three data techniques are normalized data analysis using equation "(5)" for simple (SAW), equation "(6)" for linear (VIKOR) and equation "(7)" for vector (MOORA). The calculation results are shown in "Table III".

TABLE III. PREFERENCE VALUE DATA NORMALIZED TECHNIQUE

Alts.	Simple (SAW)	Linear (VIKOR)	Vector (MOORA)
A1	0.835678	0.896219	0.504886
A2	0.715223	0.715586	0.402731
A3	0.683109	0.445157	0.341255
A4	0.886228	0.863585	0.543958
A5	0.893289	0.940866	0.550523
A↓	⋮	⋮	⋮
A297	0.633094	0.552726	0.346973
A298	0.592999	0.405014	0.299017
A299	0.715628	0.594506	0.419313
A300	0.640509	0.514882	0.342615
Min	0.478011	0.000000	0.185292
Max	0.966667	1.000000	0.606954

Preference results of data normalized simple(SAW), linear(VIKOR) and vector(MOORA) the line-chart seen in "Fig. 5".

The line-chart of normalized data preference results in "Fig. 5" shows a sample of alternative data, there are differences in the preference values of each technique, for simple techniques illustrated using the SAW method are in the value range of 0.478011 (min) and of 0.966667 (max), linear technique (VIKOR) is in the range a value of min 0 to max 1 and for vector with MOORA in the min value range 0.185292 and max 0.606954.

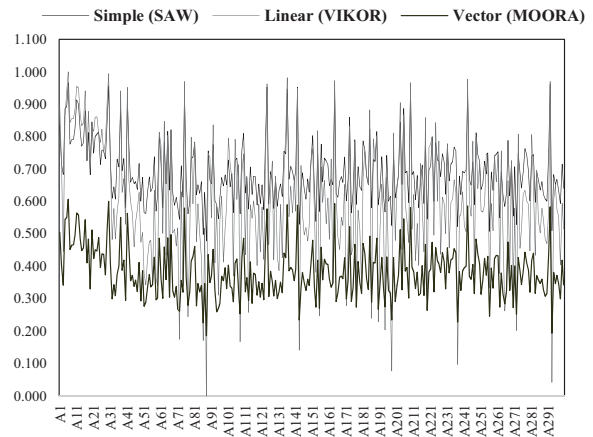


Fig. 5. Preference results of simple, linear and vector techniques

The three normalized data technique preference values show the characteristics of each method in the spread of preference values, where the simple method (SAW) tends to be centered in the spread area above the value of 0.4, linear techniques in areas above 0.5, and the vector technique centered in the distribution area above of 0.3.

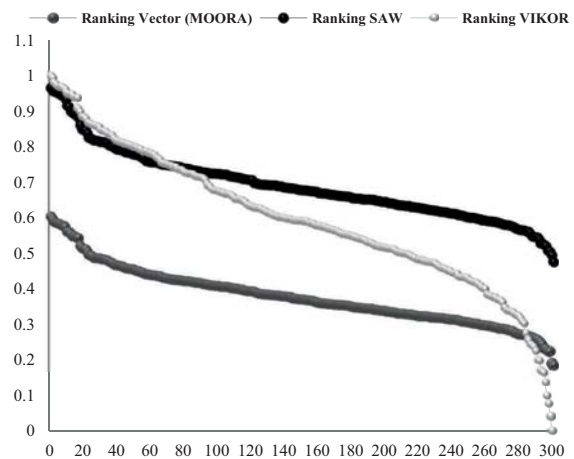


Fig. 6. Ranking results lowest to highest preference values

The preference results after being ordered from highest to lowest are obtained ranking values in each normalized data technique (see "Fig. 6), which is shown in "Table IV" of the highest alternative order.

TABLE IV. RANKING DATA NORMALIZED TECHNIQUE

Order.	Simple (SAW)	Linear (VIKOR)	Vector (MOORA)
1st	A6	A6	A6
2 nd	A292	A30	A30
3 rd	A209	A136	A164
4 th	A30	A243	A136
5 th	A142	A164	A142
A↓	⋮	⋮	⋮
297 th	A293	A237	A237
298 th	A256	A198	A86
299 th	A86	A293	A293
300 th	A88	A88	A88

The “Table IV” presents the ranking order of each technique, where there are differences in the ranking order. in the simple technique (SAW), Linear and alternative vector A6 are the highest priority, the difference is in the second order for SAW, with alternative choices A292, and the other two choose A30 and the next order experiences different positions in the alternative selection.

D. Confusion Matrix Performance Analysis

Data analysis decision table of the performance confusion matrix example for 30 target (10%) from 300 alternatives in simple(SAW) is shown in "Table V".

TABLE V. CONFUSION MATRIX FOR SIMPLE (SAW) TECHNIQUE

		Simple (SAW) Method		
		Predicted: NO	Predicted: YES	
Actual Data	Actual: NO	TPN = 270	FPP = 30	300
	Actual: YES	FPN = 3	TPP = 27	30
		273	57	

- Accuracy using(8) = 0.743 = 74.3%
- Precision using(9) = 0.474 = 47.4%
- Recall using(10) = 0.9 = 90%
- Error rate using(11) = 0.0825 = 8.25%

The performance confusion matrix example for 30 target (10%) from linear (VIKOR) is shown in "Table VI".

TABLE VI. CONFUSION MATRIX FOR LINEAR (VIKOR) TECHNIQUE

		Linear (VIKOR) Method		
		Predicted: NO	Predicted: YES	
Actual Data	Actual: NO	TPN = 270	FPP = 30	300
	Actual: YES	FPN = 7	TPP = 23	30
		277	53	

- Accuracy using “(8)” = 0.733 = 73.3%
- Precision using “(9)” = 0.434 = 43.4%
- Recall using “(10)” = 0.767 = 76.7%
- Error rate using(11) = 0.0925 = 9.25%

The performance confusion matrix example for 30 target (10%) from vector (MOORA) is shown in "Table VII".

- Accuracy using(8) = 0.740 = 74.0%
- Precision using(9) = 0.464 = 46.4%
- Recall using(10) = 0.867 = 86.7%
- Error rate using(11) = 0.085 = 8.5%

TABLE VII. CONFUSION MATRIX FOR LINEAR (VIKOR) TECHNIQUE

		Vector (MOORA) Method		
		Predicted: NO	Predicted: YES	
Actual Data	Actual: NO	TPN = 270	FPP = 30	300
	Actual: YES	FPN = 4	TPP = 26	30
		274	56	

The performance measurement data analysis of each technique is compared against the actual data. Actual data is an alternative priority target for obtaining social assistance, this actual data has attributes with the criteria value C1 and C2 in the high category (max), and C3 in the low category (min).

Furthermore, a summary of the scenario results with the percentage of the target number for the set of 10%, (30 alternatives) priority I, and the target 20% (60 alternatives) priority II is seen in "Table VIII" .

TABLE VIII. SUMMARY OF CONFUSION MATRIX PERFORMACE OF DATA NORMALIZED TECHNIQUE

Metric	Simple (SAW)		Linear (VIKOR)		Vector (MOORA)	
	30(10%)	60(20%)	30(10%)	60(20%)	30(10%)	60(20%)
Accuracy	74.3	72.3	73.3	71.3	74	71.8
Precision	47.4	45.0	43.4	42.9	46.4	43.9
Recall	0.9	81.7	76.7	75	86.7	78.3
Error rate	8.25	17.75	9.25	18.75	8.5	18.25

E. Discussion

Evaluation result of this work obtain findings during the data analysis process carried out on the 3 techniques studied in the case of social assistance data. These findings are:

- Normalized techniques Analysis in social assistance data cases, shows the characteristics for simple normalized techniques are in the distribution area of 0.4 to 0.8, for linear normalized techniques are spread over the area of 0 to1, while in normalized techniques the vector distribution of data is in the range of values of 0.3 to 0.7. Meaning that simple and linear normalized techniques have similar spread data characteristics, thus the effect on preference outcomes for social assistance decision-making cases will be the similiar also, distinct to vector normalized techniques.
- The preference results show the difference between each technique in ranking to the actual data. The simple normalization technique ranking's are better. This explains that the simple normalized technique for preference targeting 30 alternatives (small groups) is more optimal than other techniques. We observe that

there is an additive effect of the min (cost) and max (benefit) attributes on simple technique.

- The analysis results of performance using a confusion matrix show that the accuracy, precision, recall, and error rate in simple normalized techniques are better than linear and vector techniques for decisions small group targets. We observe the influence of the spread of normalized data ranges on the performance of each technique, for a certain range (with small targets) simple and linear normalization techniques will be more optimal, but when the target is increased, performance changes. The vector normalization technique performance significantly increases with the number of targets.

The findings obtained require further study, comparisons are needed with other case examples, given the limited number of criteria in the social assistance case is only a few (only 3 criteria), further research requires a wider variety of scenarios to determine the characteristics and their effects on normalized techniques, the weighting method, preference methods, etc.

V. CONCLUSION

Implementation of decision-making analysis methods in distributing internet data assistance is needed to obtain optimal results, various analytical methods can be applied and simple in the process. Each method has its own characteristics, both in preference selection and in the process of normalized data analysis. This study offers three technical approaches in normalized data analysis, simple, linear and vector techniques which are illustrated in the SAW method, VIKOR and MOORA. The selection of a method that is suitable for the case study will significantly result in optimal performance and on target.

ACKNOWLEDGMENT

Research funding comes from the Higher Education Operational Assistance Fund (BOPTN) Department of Informatics, Faculty of Engineering, Mulawarman University, Indonesia.

REFERENCES

- [1] Peraturan Pemerintah RI, *Pembatasan Sosial Berskala Besar dalam Rangka Percepatan Penanganan Corona Virus Disease 2019 (COVID-19)*. Republik Indonesia, 2020, p. PP Nomor 21 Tahun 2020.
- [2] A. I. Fauzi Widyaiswara Ahli Madya and B. Pengembangan Sumber Daya Manusia Provinsi Riau Jalan Ronggowarsito nomor, "Implementasi Pembatasan Sosial Berskala Besar, Sebuah Kebijakan Publik Dalam Penanganan Pandemi Covid-19," 2020.
- [3] Pusdiklat Kemdikbud, "Surat Edaran MENDIKBUD No 4 Tahun 2020." <https://jdih.kemdikbud.go.id>.
- [4] D. Pavlicic, "Normalisation affects the results of MADM methods," *Yugoslav Journal of Operations Research*, 2001.
- [5] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, "Data normalisation techniques in decision making: Case study with TOPSIS method," *International Journal of Information and Decision Sciences*. 2018, doi: 10.1504/IJIDS.2018.090667.
- [6] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, "Selection of normalization technique for weighted average multi-criteria decision making," 2018, doi: 10.1007/978-3-319-78574-5_4.
- [7] E. Budiman, N. Dengen, Haviluddin, and W. Indrawan, "Integrated multi criteria decision making for a destitute problem," in *Proceeding - 2017 3rd International Conference on Science in Information Technology, ICSITech 2017*, 2017, vol. 2018-Janua, pp. 342–347, doi: 10.1109/ICSITech.2017.8257136.
- [8] A. Jahan and K. L. Edwards, "A state-of-the-art survey on the influence of normalization techniques in ranking: Improving the materials selection process in engineering design," *Materials and Design*. 2015, doi: 10.1016/j.matdes.2014.09.022.
- [9] P. Chatterjee and S. Chakraborty, "Investigating the effect of normalization norms in flexible manufacturing system selection using multi-criteria decision-making methods," *Journal of Engineering Science and Technology Review*, 2014, doi: 10.25103/jestr.073.23.
- [10] E. Budiman, "Mobile Data Usage on Online Learning during COVID-19 Pandemic in Higher Education," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 14, no. 9, 2020.
- [11] M. Danielson and L. Ekenberg, "Trade-offs for ordinal ranking methods in multi criteria decisions," 2017, doi: 10.1007/978-3-319-52624-9_2.
- [12] E. Budiman, "Importance-weighted Ranking Methods for Preference the Covid-19 Pandemic Social Assistance," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 10, no. 1, pp. 108–115, 2020, doi: 10.35940/ijeat.A1743.1010120.
- [13] M. Wati, N. Novirasari, E. Budiman, and Haeruddin, "Multi-Criteria Decision-Making for Evaluation of Student Academic Performance Based on Objective Weights," in *2018 Third International Conference on Informatics and Computing (ICIC)*, 2018, pp. 1–5, doi: 10.1109/IAC.2018.8780421.
- [14] A. Jahan and K. L. Edwards, "VIKOR method for material selection problems with interval numbers and target-based criteria," *Materials and Design*, 2013, doi: 10.1016/j.matdes.2012.12.072.
- [15] W. K. Brauers and E. K. Zavadskas, "Robustness of the multi-objective moora method with a test for the facilities sector," *Technological and Economic Development of Economy*, 2009, doi: 10.3846/1392-8619.2009.15.352-375.
- [16] E. Budiman, "Decision Optimization: Internet Data Assistance for Students during Learning from Home," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 11, pp. 372–378, 2020, doi: 10.35940/ijitee.K7845.0991120.
- [17] E. Budiman, A. Lawi, and S. La Wungo, "Implementation of SVM Kernels for Identifying Irregularities Usage of Smart Electric Voucher," 2020, doi: 10.1109/iccdd46541.2019.9161077.
- [18] N. Dengen, Haviluddin, L. Andriyani, M. Wati, E. Budiman, and F. Alameka, "Medicine Stock Forecasting Using Least Square Method," 2018, doi: 10.1109/EICConCIT.2018.8878563.
- [19] E. Budiman, Haviluddin, N. Dengan, A. H. Kridalaksana, M. Wati, and Purnawansyah, *Performance of Decision Tree C4.5 Algorithm in Student Academic Evaluation*, vol. 488. 2018.