

COMPARISON OF LINEAR AND VECTOR DATA NORMALIZATION TECHNIQUES IN DECISION MAKING FOR LEARNING QUOTA ASSISTANCE

Edy Budiman¹, Ummul Hairah²

^{1,2}Dept. of Informatics Faculty of Engineering Mulawarman University, East-Kalimantan Indonesia
¹edy.budiman@fkti.unmul.ac.id

Abstract-- Data normalization is essential for all kinds of decision-making problems, and a lot of effort has been spent on the development of normalization models in multi-criteria decision making (MCDM), but despite all this, there is no definite answer to the question: Which is the most appropriate technique?. This paper compares the popular normalization techniques: Linear Normalization (LN) and Vector Normalization (VN) using VIšekriterijumsko KOMPromisno Rangiranje (VIKOR) Method. The beneficiaries dataset of learning quota was collected of 399 students sample through observation (drive-test measurements and online questionnaires) to obtain information on criteria data including attributes in online learning during the Covid-19 pandemic. The ranking results for vector vs linear normalization show how ranking is affected. The difference in the selection of the best alternative (rank) shows that there are differences in vector and linear assessments that are influenced by the max-min criterion value which has an impact on the rank-sum results (benefit/cost). This test clearly shows how important it is to use an appropriate (normalized) representation of the model because there will often be a criterion where "the higher the better" while for others (cost) "the lower the better".

Keywords: Normalization; Linear; Vector; Decision-making; Learning-Quota.

I. INTRODUCTION

In most multi-criteria decision-making (MCDM) problems, the criteria are of different scales (eg packet data consumption, study load, economic capacity, costs, etc. in the case of learning quota assistance). Thus, we have to use multiple pre-processing to obtain the same scale, which will allow the aggregation of numerical and comparator criteria to obtain a final score for each alternative. In general, the MCDM model consists of the alternative set A_i ($i = 1, \dots, m$), the criteria set C_j (j

$= 1, \dots, n$) and the appropriate criteria weight W_j . Furthermore, r_{ij} is the value of the decision matrix, which classifies Alternative i concerning the criterion j . Normalizing the value of the decision matrix, r_{ij} , we obtain a dimensionless element, which can be combined to obtain a rank per alternative [1].

However, there have been very few studies on normalization techniques and how to select suitable techniques for MCDM problems, and this is the motivation for this article. It should also be noted that if the normalization technique is not suitable for the decision problem or the method chosen, the best decision solution may be missed. As Chatterjee and Chakraborty state that "In fact, while the normalization process scales the criterion value to approximately the same magnitude, different normalization techniques can produce different solutions and, therefore, can cause deviations from the original recommended solution. [2]. In addition, in the current era of the internet of things and as a result, large amounts of data are available, the question of appropriate normalization techniques poses a bigger challenge, as there will be an explosion of criteria and alternatives and scaling them into dimensionally more difficult units.

In general, data normalization in decision-making analysis is a transformation process to obtain numerical input data and its comparison using the same scale [3]. The normalization technique maps attributes (criteria) with different units of measurement to the same scale in the interval: 0-1.

The Multi-criteria Decision Making (MCDM) method can determine how to attribute information

is processed to arrive at a choice, requires comparisons between and between attributes, and involves appropriate explicit exchanges[4].

Several studies on normalization matrices such as that of Chuan Yue normalize attributes for group decision making and applications for software reliability assessment[5]. S. H. Zolfani focused on re-analysis of the MADM method based on logarithmic normalization[6]. Nazanin Vafaei[7] Data normalization techniques in decision making using the TOPSIS method. examined the effect of normalization techniques on ranking: Improving the material selection process in engineering design[8]. Aydin Çelen analyzed the comparison of normalization procedures in the TOPSIS method: By application to the Turkish deposit banking market [9], and others.

Summarizing the first challenge of modeling and applying the MCDM method to solving decision problems is selecting the appropriate normalization technique for the problem at hand. In this study, we discuss the effect of applying Linear and Vector normalization techniques based on optimization of the Sum-Based ratio analysis of benefit-cost criteria in the case of internet data quota assistance from the Ministry of Education and Culture (KEMENDIKBUD) to support online learning of students during the Covid-19 pandemic.

II. METHODOLOGY

The Indonesia Ministry of Education and Culture (KEMENDIKBUD) has issued an internet data quota assistance program since 2020[10], the latest technical guidelines for this policy are in the Secretary-General Regulation Number 4 of 2021 concerning Technical Instructions for Distribution of Government Assistance for Internet Data Quota Packages for 2021 [11]. The form of assistance provided is internet data quota with details of the amount of Early Childhood Education (PAUD) assistance of 7GB/month, Primary and Secondary Education 10GB/month, Students and Lecturers of 15GB/month and for Educators of 12GB/month for 3 months. Based on the Assistance Pocket Book, the internet data quota package has a validity period of 30 days from the time the internet data quota package is received by the mobile number of

educators and students, and the remaining internet data package quota that is not used every month will expire or is not cumulative for the following month [12].

The internet data quota assistance policy to support online learning from home is the right program during the Covid-19 pandemic. However, it will be more optimal and right on target if the decision-making involves stakeholders and the distribution is based on the criteria for learning needs and the economic capacity of potential beneficiaries. Given the needs of each student are different from each other, differences in learning load, economic ability, duration of online meetings, and others. Optimizing the distribution of internet data quota package assistance through the application of criteria and data normalization techniques for decision making based on the needs of potential beneficiaries so that they are right on target and objectively or proportionally.

A. Design Model

An overview of the design model of applying Linear and Vector normalization techniques in decision-makings on internet data quota assistance is presented in “Fig. 1”.

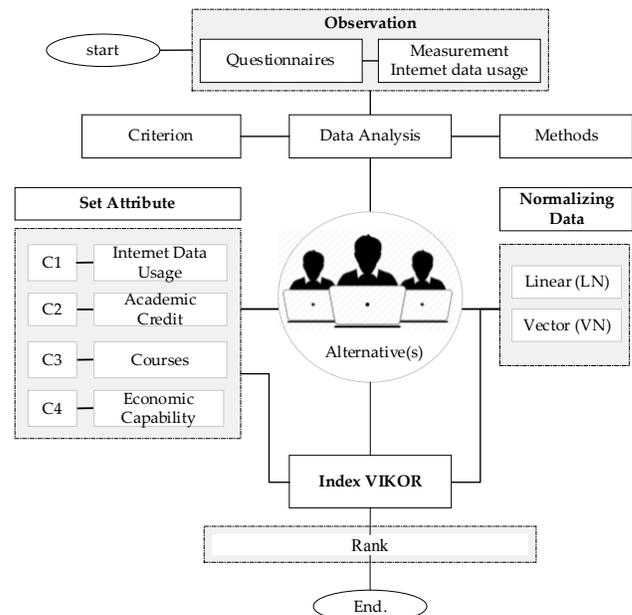


Fig. 1 Design Model Linear and Vector in Decision-making

B. Data Collection Methods

The research samples for internet data quota

assistance were students and lecturers of the Department of Informatics, Mulawarman University Samarinda, East Kalimantan. Data collection methods for internet data quota assistance criteria use observation through drive-test measurements to obtain information on the amount of student internet data usage in online meetings [13]. Data collection and other criteria information using online questionnaires distributed using Google forms.

C. Internet Data Quota Assistance Criteria

The decision-making criteria for students' internet data quota assistance to support online learning from home uses 4 criteria obtained from the results of internet data measurement with Zoom-Meeting and online questionnaires. Data criteria are shown in “Table 1”.

TABLE I
 Attribute and Criterion

Code	Criteria	Definitions	Attribute
C1	Internet Data Usage	Measurements result from internet data usage students via Zoom meetings (MB)	Benefit
C2	Academic Credits	Number of Student academic credits	Benefit
C3	Courses	Number of Student Courses	Benefit
C4	Economic Capability	Student economic capability per month (IDR)	Cost

D. Data Normalization Technique

Most MCDM models require a normalization stage, which is determined by a decision matrix that has the following parts: alternatives A_i ($i = 1, \dots, M$), criterion C_j ($j = 1, \dots, n$), criterion relative importance (or weights) w_j , and the decision matrix with the element r_{ij} , which is the alternative rating i associated with criterion j [14].

The flowchart of the comparison process between the two methods is simply shown in “Fig. 2.

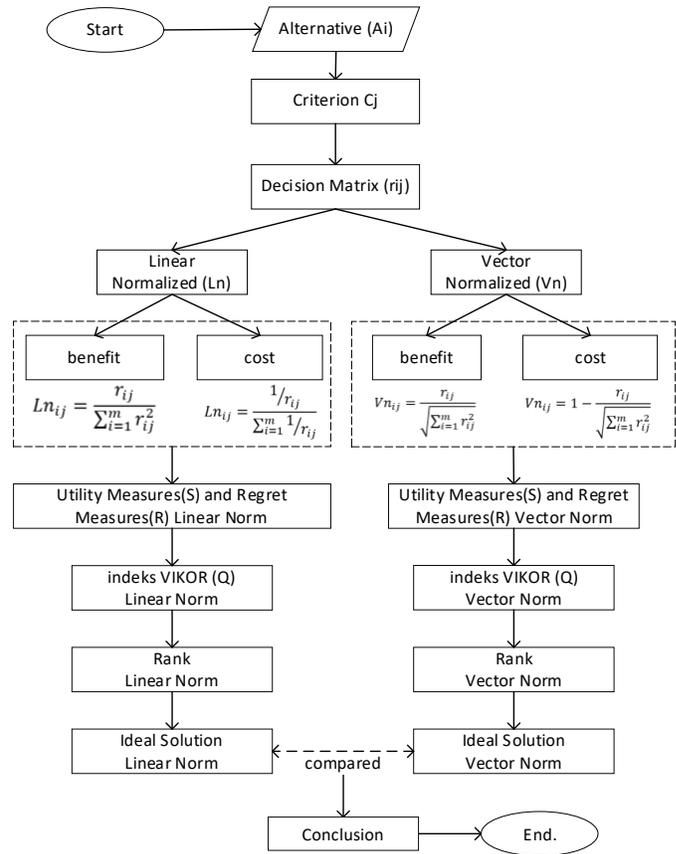


Fig. 2. Flowchart comparison of linear and vector methods

Fig. 2 shows the comparison process flowchart of linear and vector normalization methods. The basic difference is in calculating the attribute value (benefits and costs) in normalizing r_{ij} data. Attributes for which a higher value is desired are called positive criteria or benefit and property attributes with a smaller value, are named negative criteria, cost criteria, or non-benefit attributes. The data normalization results (linear and vector) were then analyzed using the VIšekriterijumsko KOMpromisno Rangiranje (VIKOR) method to find the utility (S), Regret (R) measures, and the VIKOR index. The results of sorting the value (rank) of the two methods are the ideal solution for decision-making.

Table II shows the Sum-based methods available for the benefit and cost criteria.

TABLE II
 Data Normalization Linear dan Vector Technique

Norm	Attribute	Equations
Vector	Benefit	$Vn_{ij} = \frac{r_{ij}}{\sqrt{\sum_{i=1}^m r_{ij}^2}} \quad (1)$
	Cost	$Vn_{ij} = 1 - \frac{r_{ij}}{\sqrt{\sum_{i=1}^m r_{ij}^2}} \quad (2)$
Linear	Benefit	$Ln_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}^2} \quad (3)$
	Cost	$Ln_{ij} = \frac{1/r_{ij}}{\sum_{i=1}^m 1/r_{ij}} \quad (4)$

III. RESULTS AND DISCUSSION

The data quota assistance dataset is collected through observations (drive-test measurements and online questionnaires) to obtain data information on criteria including attributes. The population of this study was 1000 students of the Informatics department at Mulawarman University in online learning during the Covid-19 pandemic. Determination of the sample using reference tables Isaac and Michael [15] with error rate alpha 1% (0.01), obtains a sample of 399 students (male 213, female 186). Descriptives data are shown in Table III.

TABLE III
 Descriptives Dataset

Alt.	C1	C2	C3	C4
A1	922.84	23	8	922.84
A2	687.62	17	6	687.62
A3	846.39	20	7	846.39
A4	998.73	21	8	998.73
A5	1002.54	24	9	1002.54
⋮	⋮	⋮	⋮	⋮
A396	1043.49	24	8	3600000
A397	871.57	20	7	1750000
A398	787.22	19	7	2450000
A399	1002.31	23	8	1200000
min	455.78	14	5	900000
max	1129.97	24	9	3850000
mean	766.45	20.27	7.04	2140902.26
mode		21	7.00	2100000

Table III presents a description of student internet data quota assistance with criteria C1, C2, and C3 with benefit attributes and C4 criteria with cost attributes, meaning that the top priority for potential beneficiaries is those with the use of internet data, the max number of academic credits and courses but low economic capacity. Meanwhile, the categories that are not a priority for assistance are those who have min internet data usage, min number of credits and courses, and have high economic capacity.

A. Results: Vector Normalization Analysis

Data Normalization of vector that the denominator is the square root of the sum of the squares of each alternative per attribute. The vector normalization calculation use “(1)” for the benefit attribute criteria (C1, C2, and C3) and uses “(2)” for the cost attribute criteria (C4). The calculation results obtained that the value of each alternative is presented in “Table IV” and “Fig. 3”.

TABLE IV
 Results: Data Normalization for Vector Technique

Alt.	(C1) benefit	(C2) benefit	(C3) benefit	(C4) cost
A1	0.0591	0.0563	0.0563	0.9324
A2	0.0440	0.0416	0.0423	0.9437
A3	0.0542	0.0489	0.0493	0.9516
A4	0.0640	0.0514	0.0563	0.9178
A5	0.0642	0.0587	0.0634	0.9516
A↓	⋮	⋮	⋮	⋮
A397	0.0504	0.0465	0.0493	0.9448
A398	0.0642	0.0563	0.0563	0.9730
min	0.0292	0.0343	0.0352	0.9133
max	0.0724	0.0587	0.0634	0.9797

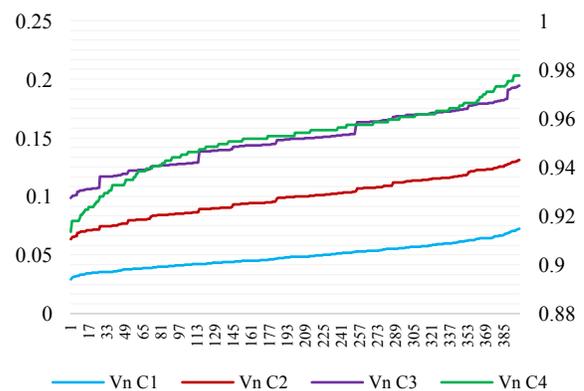


Fig 3. Line Graph Vector Norm Criteria Values

Table IV and “Fig. 3” presents the calculation results of the Vector Normalization (Vn) value for each criterion, the value range min-max is 0.0292 - 0.0724 for C1, C2 (0.0343 - 0.0587), C3 (0.0352 - 0.0634) and for C4 range (0.9133 - 0.9797).

B. Results: Linear Normalization Analysis

Normalized data analysis for the Linear technique use “(3)” for the criteria with benefit attributes (C1, C2, and C3) and use “(4)” for the cost attribute criteria (C4). The calculation results obtained the value of each alternative (A1 to A399). Head and Tail Linear Normalization (LN) results are presented in “Table V”.

TABLE V
Results: Data Normalization for Linear Technique

Alt.	(C1)	(C2)	(C3)	(C4)
	benefit	benefit	benefit	cost
A1	0.0030	0.0028	0.0028	0.0017
A2	0.0022	0.0021	0.0021	0.0020
A3	0.0028	0.0025	0.0025	0.0023
A4	0.0033	0.0026	0.0028	0.0014
A5	0.0033	0.0030	0.0032	0.0023
A↓	⋮	⋮	⋮	⋮
A398	0.0026	0.0023	0.0025	0.0020
A399	0.0033	0.0028	0.0028	0.0042
min	0.0015	0.0017	0.0018	0.0013
max	0.0037	0.0030	0.0032	0.0050

Table V the calculation results of the LN value for each criterion, obtained a value range min-max of 0.0015 - 0.0037 for C1, C2 (0.0017 - 0.0032), C3 (0.0018 - 0.0032) and C4 (0.0013 - 0.0050). The graph of the LN value is presented in “Fig. 4”.

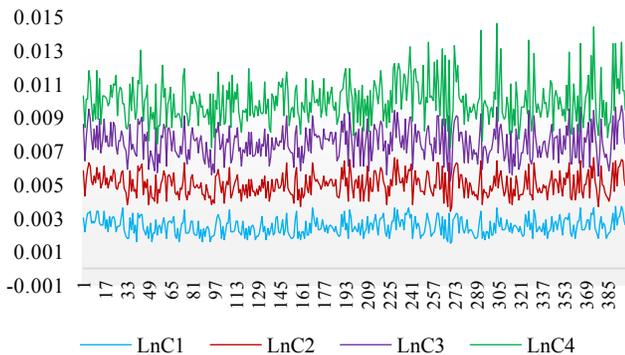


Fig 4. Line Graph Linear Norm Criteria Values

C. Results: Benefit+Cost (Sum-based)

Calculates the Sum-based value, where the normalized size is added up for the benefit attribute (C1 + C2 + C3) and subtracted for the cost attribute (C4) or subtracts the benefit-cost value for each row to get the ranking on each row. Sum-based calculation results for vector and linear normalization are presented in "Table VI".

TABLE VI
Results: Sum-Based for Data Norm: Vector and Linear

Alt	Vektor			Linear		
	benefit	cost	sum	benefit	cost	sum
A1	0.1717	0.932	1.104	0.0087	0.0017	0.010405
A2	0.1279	0.943	1.071	0.0064	0.0020	0.008452
A3	0.1524	0.951	1.104	0.0077	0.0023	0.010013
A4	0.1717	0.918	1.089	0.0084	0.0014	0.009859
A5	0.1863	0.951	1.138	0.0092	0.0023	0.011513
A↓	⋮	⋮	⋮	⋮	⋮	⋮
A398	0.1462	0.945	1.091	0.0072	0.0020	0.00927
A399	0.1768	0.973	1.149	0.0089	0.0042	0.01317
min	0.1011	0.913	1.036	0.0052	0.0013	0.00717
max	0.1921	0.977	1.167	0.0095	0.0050	0.01451

The result of the calculation is the alternative that has the highest final value so that alternative is the best alternative from the existing data, this alternative will be chosen according to the existing problems because this is the best choice. Meanwhile, the alternative that has the lowest final value is the worst alternative from existing data.

TABLE VII
Tabel 7. Ranking Normalisasi Vektor vs Linear

Rank	Vektor (Vn)		Linear (Ln)	
	Alt	Sum-based	Alt	Sum-based
1 st	A304	1.1679	A375	0.014508
2 nd	A395	1.1629	A292	0.014340
3 rd	A375	1.1625	A304	0.014296
4 th	A365	1.1616	A254	0.013633
5 th	A391	1.1603	A389	0.013507
6 th	A327	1.1598	A273	0.013387
7 th	A292	1.1592	A327	0.013309
8 th	A357	1.1542	A264	0.013255

Comparison results from two normalization techniques are shown in “Fig. 5”. As expected, the ranking of alternatives differs when using different normalization techniques.

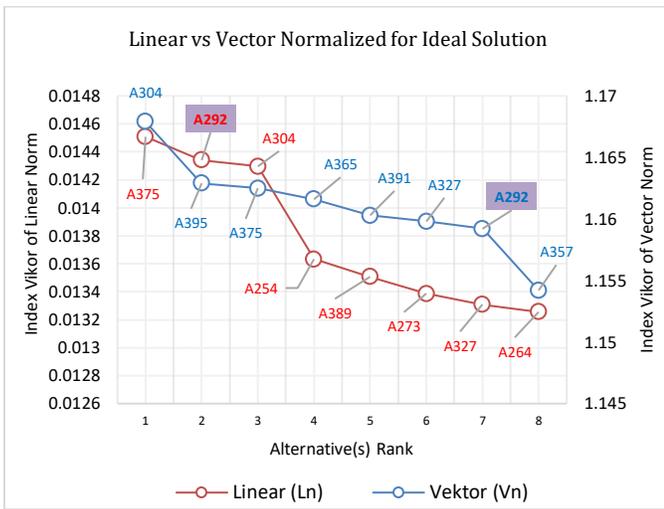


Fig. 5. Top 8 Linear vs Vector Normalized for decision ideal solution

Fig. 5 shows the difference in the decision of the best alternative ideal solution from the two methods. The highlighted alternative appears in alternative 292, where the linear method sets A292 in the 2nd rank, while the status of A292 in the vector method is in the 7th rank (top 8).

The best rank (ideal solution) for prospective beneficiaries is prioritized for students who have a learning load in the very high category but have very low economic capability. Refer to the actual data, Alternative 292 has an attribute value of Internet Data Usage (C1) of 1038.08 MB (Very-high), Academic Credits (C2) of 18 credits (high), C3 of 6 courses (very low), and economic capability of IDR 1,000,000 (Very low). The value of A292 is obtained from the addition of the benefit attribute, the reduction for the cost attribute (C1 + C2 + C3) – (C4). The difference in the value of the calculation results causes the assessment of the ranking order of each method to be different which is influenced by the magnitude of the value of the benefit-cost attributes.

D. Discussion

Normalization provides a way to compare all attributes on the same scale, thus allowing the evaluation of each alternative with one value, and then the choice of the best alternative according to the size obtained by the applied MCDM method. Based on the results of the analysis of the application of 2 normalization methods (Vector and Linear) in the case of internet quota assistance,

several points of findings are explained:

The use of normalization techniques in the case of internet data quota assistance can handle the problem of making decisions on several distinctive attributes. The ability to remove scales is a basic rule that when normalizing identical data with different units or scales, the same result is obtained. The function of the same criteria can be demonstrated using different 'conversion' units, for example, internet data consumption in Megabytes (MB), Academic credit, Courses, or Economic capability in IDR.

This 'conversion' unit affects the grading, and the application of the normalization method can eliminate the unit function of the 'changeable' criterion, and returns the same result for all scales.

To keep the maximum initial information concerning the initial attribute values and other criterion values, it is necessary to check the symmetry of the normalized values when comparing the cost and benefit criteria. For example, benefit criteria data can be normalized in the p-1 interval (0<p<1), whereas for cost-type criteria the value is included in the interval from 0-p or 0-1. This can be a type of asymmetry that some methods cannot cover the entire range of 0 and 1. Other types of asymmetry appear at normal values of the same data when normalized as a cost and benefit criterion. Linear and vector normalize only eliminates the criteria scale but cannot convert the cost criterion into a benefit criterion.

Ranking results for vector vs linear normalization in “Fig. 5”, we see how ranking is affected. For example, A292 is the best ranking alternative to vector normalization while A292 is the best alternative to linear normalization. Reviewing A304 data properties with data usage requirements (C1) 1,103.52MB per day, 23 credits (C2), 9 courses (C3) and has an economic capacity of IDR 1 million / month). For the A375 data property, the use of data is 1,092.91MB, 24 credits, 8 courses, and 1 million / month economic capacity. Based on the property of these two alternatives, it shows the difference in vector and linear assessment which is influenced by the max-min criterion value which has an impact on the rank-sum result (benefit-cost).

This test clearly shows how important it is to use an appropriate (normalized) representation of the model as there will often be a criterion where "the higher the better" while for others (cost) "lower is better".

IV. CONCLUSION

Normalization is an inseparable part of the decision-making process because we need to get dimensionless units to calculate the final rank per alternative. This study in the case of internet data quota assistance shows some of the effects of using different amount-based normalization techniques and also the importance of distinguishing between benefit and cost criteria.

For this effect, we compared the technique - vector and linear normalization. The results of the comparison with the two normalization techniques, as expected, rank the alternatives differently when using different normalization techniques. It is interesting to see that there is complete consensus on the difference between the best alternatives and all the other rankings being relatively different, therefore another evaluation method is needed to judge which technique is best to apply to the context of this problem. Because it is difficult to judge which normalization technique is best just by looking at the results obtained. This is a preliminary study, which we plan to extend to other normalization techniques that may prove more adequate for MCDM problems. Plus, we'll look at how it affects big data in today's internet age.

V. ACKNOWLEDGMENT

Thanks to the Dean of the Faculty of Engineering, Study Program Coordinator, and Informatics Student, Mulawarman University, Samarinda, East Kalimantan for their support and participation in the research.

VI. REFERENCES

- [1] S. Panjwani, S. N. Kumar, and L. Ahuja, "Multi-criteria decision making and its applications," *Int. J. Innov. Technol. Explor. Eng.*, 2019, doi: 10.35940/ijitee.II122.0789S419.
- [2] P. Chatterjee and S. Chakraborty, "Investigating the effect of normalization norms in flexible manufacturing system selection using multi-criteria decision-making methods," *J. Eng. Sci. Technol. Rev.*, 2014, doi: 10.25103/jestr.073.23.
- [3] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, "Normalization techniques for multi-criteria decision making: Analytical hierarchy process case study," 2016, doi: 10.1007/978-3-319-31165-4_26.
- [4] A. B. Pandey, P. B. Nawale, R. R. Patil, S. S. Patil, and V. B. Sawant, "Effect of Normalization Techniques on MEMS Digital Micromirror Selection using MADM Methodology," *Int. J. Res. Eng. Appl. Manag.*, no. 01, pp. 67–75, 2020, doi: 10.35291/2454-9150.2020.0259.
- [5] C. Yue, "Attribute Normalization Approaches to Group Decision-making and Application to Software Reliability Assessment," *Cognit. Comput.*, 2021, doi: 10.1007/s12559-019-09707-2.
- [6] S. H. Zolfani, M. Yazdani, D. Pamucar, and P. Zarate, "A vikor and topsis focused reanalysis of the madm methods based on logarithmic normalization," *Facta Univ. Ser. Mech. Eng.*, 2020, doi: 10.22190/FUME191129016Z.
- [7] 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.
- [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] A. Çelen, "Comparative analysis of normalization procedures in TOPSIS method: With an application to Turkish deposit banking market," *Inform.*, 2014, doi: 10.15388/Informatica.2014.10.
- [10] Kemendikbud, "Peraturan Sekretaris Jenderal Nomor 14 Tahun 2020, tentang Petunjuk Teknis Bantuan Kuota Data Internet Tahun 2020," Jakarta, 2020. [Online]. Available: <https://kuota-belajar.kemdikbud.go.id>.
- [11] Kemendikbud, "Peraturan Sekretaris Jenderal Nomor 4 Tahun 2021 Tentang Petunjuk Teknis Penyaluran Bantuan Pemerintah Paket Kuota Data Internet Tahun 2021." 2021, [Online]. Available: <https://kuota-belajar.kemdikbud.go.id/>.
- [12] Kemendikbud, "Buku Saku Bantuan Kuota Data Internet 2021." Kemendikbud, 2021, [Online]. Available: <https://kuota-belajar.kemdikbud.go.id>.
- [13] E. Budiman, "Mobile Data Usage on Online Learning during COVID-19 Pandemic in Higher Education," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 19, pp. 4–17, 2020, doi: 10.3991/ijim.v14i19.17499.
- [14] E. Budiman, N. Dengen, Haviluddin, and W. Indrawan, "Integrated multi criteria decision making for a destitute problem," in *2017 3rd International Conference on Science in Information Technology (ICSITech)*, 2017, pp. 342–347, doi: 10.1109/ICSITech.2017.8257136.
- [15] P. D. Sugiono, "Metode penelitian pendidikan pendekatan kuantitatif.pdf," *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif Dan R&D*. 2014.