JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 1, Januari 2021, Page 42-47
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2523

Optimizing the Division of Study Class Groups Using the Partitioning Around Medoids (PAM) Method

Frainskoy Rio Naibaho

Prodi Pendidikan Agama Kristen, Institut Agama Kristen Negeri Tarutung, Tarutung, Indonesia
Email: frainskoy.rio.naibaho@email.com
Email Penulis Korespondensi: frainskoy.rio.naibaho@gmail.com

Abstract−Optimization is a step to solve a problem to get more profitable results. Profitable based on the point of view used or the desired needs. The optimization value can be profitable in the maximum position or profitable in the minimum position. A problem can be solved in different ways, to produce the best solution. The best conditions can be viewed from many things, including tolerance, methods, and problems. Many theories have been developed to solve optimization problems. This optimization problem is often discussed because it is very close to human life. In this case, optimization can be interpreted as the process of achieving the most optimal results by adjusting input, selecting equipment, mathematical processes, and testing. Thereby in this paper, the Partitioning Around Medoids (PAM) method has succeeded in optimizing class grouping by calculating the closest distance between the achievement and intelligence of each student.

Keywords: Cluster; Partitioning Around Medoids; Optimization; Grouping; Decision Support System

1. INTRODUCTION

Class is another form of set or grouping. Each class is filled with a variety of different variables. In the formation of a class ideally, class members can be grouped based on uniformity of achievement [1]. This uniformity of achievement can affect the teaching process more effectively and efficiently. The level of intelligence of students is very influential in the achievement of material in the classroom [2]. Because students with above-average intelligence will prefer a fast learning process and material that is always increasing. Whereas students with lower levels of intelligence would prefer a slower and more repetitive learning process [3]. With this difference, the learning process will experience inequality [4].

The correct teaching process can also improve the quality of learning. This class grouping needs to be done because there are differences in variations between the abilities and skills of each student [5], [6]. Schools often classify students using a random method, ignoring student achievement and intelligence levels [7]. The right method in group formation is to use the cluster method. The cluster method can group students mathematically and can reduce the boundaries between students and other students. Based on this explanation, it is necessary to research grouping student classes using the PAM cluster technique.

The PAM algorithm is better than other algorithms because it uses the entire dataset to find the best potential. This algorithm is classified as an algorithm with high computational complexity because it always calculates the distance between datasets[8][9]. The PAM method is also proven to be stronger than the k-means method in clustering because k-means has a weakness in handling noise and outliners[10], [11].

2. RESEARCH METHODOLOGY

2.1 Clustering

Clustering is a method of analyzing data, which is often included as a data mining method. The purpose of clustering is to group data with the same characteristics into the same area and data with different characteristics to another region[12]. Various methods can be used to measure the similarity value between objects being compared, one of which is the Euclidean Distance[13].

$$dist(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

2.2 Optimization

Optimization aims to be able to solve a certain problem so that the most favorable conditions are obtained from a certain point of view [14]. Optimization is a process related to input adjustment, characteristic selection, mathematical process, and testing [15].

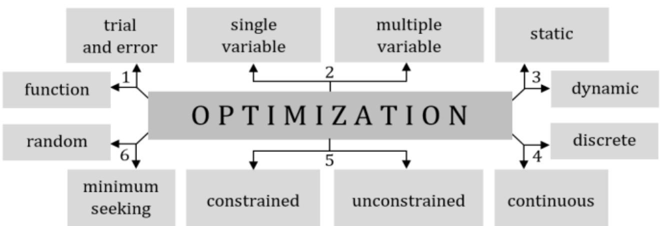

Figure 1. Optimization classification

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 1, Januari 2021, Page 42-47
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2523

2.3 Partitioning Around Medoids (PAM)

K-Medoid or also known as Partitioning Around Medoids (PAM) is a classic partition clustering technique that groups data sets from n_i objects into k groups known as a priori. The cluster formation process begins by randomly assigning k objects from the dataset as medoids[9].

Step of PAM algorithm:

a. Initialize k cluster centers (*number of clusters*)
b. Allocate each data (*object*) to the nearest cluster using the Euclidian Distance measurement equation
c. Randomly select objects in each cluster as candidates for the new medoid.
d. Calculate the distance of each object in each cluster with the new medoid candidate.
e. Calculate the total deviation (S) by calculating the value of the new total distance - the old total distance. If S <0, then swap objects with cluster data to form a set of k new objects as medoids.
f. Repeat steps 3 to 5 until there is no change in medoid, so that you get the cluster and the members of each cluster.

Tabel 1. Algoritma K-Medoids

```
Algorithm 1: K-Medoids process
function MAP(dataset)
 initialize medoid gets cluster center
  Euclidean Distance of dataset
   Calculate Total distance
  If (new of Total distance – Total distance) < 0
   Update initial cluster
 clustering
end function
```

3. RESULT AND DISCUSSION

The initial stage is to prepare a dataset. In this study using data on school X with the average school score and the average exam score shown in table 1. The average school score was obtained from the diploma, while the average exam score was obtained at the time of school entrance selection. The data used are as many as 40 data, which will be grouped into two clusters. The data grouping is used in the following stages:

Tabel 2. Datasets

No	Name	Average value		No	Name	Average value	
		From	Exam			From	Exam
1	P1	88	68	21	P21	85	94
2	P2	85	95	22	P22	87	84
3	P3	87	73	23	P23	84	90
4	P4	83	81	24	P24	85	75
5	P5	85	80	25	P25	89	66
6	P6	84	80	26	P26	87	88
7	P7	84	62	27	P27	81	60
8	P8	83	61	28	P28	88	89
9	P9	89	68	29	P29	84	64
10	P10	90	92	30	P30	91	70
11	P11	86	92	31	P31	86	61
12	P12	85	95	32	P32	89	73
13	P13	85	90	33	P33	89	87
14	P14	80	81	34	P34	82	70
15	P15	87	63	35	P35	86	90
16	P16	84	85	36	P36	85	97
17	P17	89	73	37	P37	87	70
18	P18	83	84	38	P38	86	85
19	P19	82	71	39	P39	86	71
20	P20	85	93	40	P40	86	93

participant P = {1, 2, ..., 40}

Initialization of the cluster center as much as 2 clusters from the sample data. For selection, each medoid was selected randomly. Assume P5 and P17 as initial medoids P5(85, 80), P17(89, 73).

Submitted: **02/10/2020**; Accepted: **19/11/2020**; Published: **22/01/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 1, Januari 2021, Page 42-47
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2523

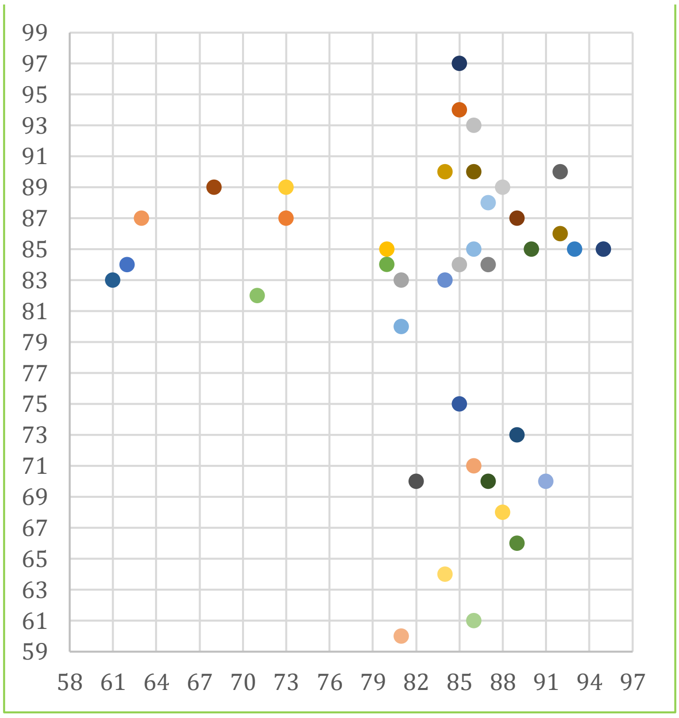

Figure 2. Dataset distribution

Calculate the distance of the object to each medoid that has been selected using the distance formula that is most often used, namely the Euclidean distance.

C1

$d(P1, P5) = \sqrt{(88-85)^2 + (68-80)^2} = 12.369$

$d(P2, P5) = \sqrt{(85-85)^2 + (95-80)^2} = 15$

$d(P3, P5) = \sqrt{(87-85)^2 + (73-80)^2} = 7.28$

$d(P5, P5) = \sqrt{(85-85)^2 + (80-80)^2} = 0$

$d(P38, P5) = \sqrt{(86-85)^2 + (85-80)^2} = 5.099$

$d(P39, P5) = \sqrt{(86-85)^2 + (71-80)^2} = 9.055$

$d(P40, P5) = \sqrt{(86-85)^2 + (93-80)^2} = 13.038$

C2

$d(P1, P17) = \sqrt{(88-89)^2 + (68-73)^2} = 5.099$

$d(P2, P17) = \sqrt{(85-89)^2 + (95-73)^2} = 22.361$

$d(P3, P17) = \sqrt{(87-89)^2 + (73-73)^2} = 2$

$d(P5, P17) = \sqrt{(86-89)^2 + (93-73)^2} = 20.224$

$d(P38, P17) = \sqrt{(86-89)^2 + (85-73)^2} = 12.369$

$d(P39, P17) = \sqrt{(86-89)^2 + (71-73)^2} = 3.606$

$d(P40, P17) = \sqrt{(86-89)^2 + (93-73)^2} = 20.224$

The overall results of the first calculation can be seen in Table 3, as follows:

Tabel 3. Datasets first iteration

No	Name	C1	C2	Minimum distance	cluster	No	Name	C1	C2	Minimum distance	cluster
1	P1	12.369	5.099	5.099	2	21	P21	14.000	21.378	14.000	1
2	P2	15.000	22.361	15.000	1	22	P22	4.472	11.180	4.472	1
3	P3	7.280	2.000	2.000	2	23	P23	10.050	17.720	10.050	1

Volume 5, Nomor 1, Januari 2021, Page 42-47
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2523

4	P4	2.236	10.000	2.236	1	24	P24	5.000	4.472	4.472	2
5	P5	0.000	8.062	0.000	1	25	P25	14.560	7.000	7.000	2
6	P6	1.000	8.602	1.000	1	26	P26	8.246	15.133	8.246	1
7	P7	18.028	12.083	12.083	2	27	P27	20.396	15.264	15.264	2
8	P8	19.105	13.416	13.416	2	28	P28	9.487	16.031	9.487	1
9	P9	12.649	5.000	5.000	2	29	P29	16.031	10.296	10.296	2
10	P10	13.000	19.026	13.000	1	30	P30	11.662	3.606	3.606	2
11	P11	12.042	19.235	12.042	1	31	P31	19.026	12.369	12.369	2
12	P12	15.000	22.361	15.000	1	32	P32	8.062	0.000	0.000	2
13	P13	10.000	17.464	10.000	1	33	P33	8.062	14.000	8.062	1
14	P14	5.099	12.042	5.099	1	34	P34	10.440	7.616	7.616	2
15	P15	17.117	10.198	10.198	2	35	P35	10.050	17.263	10.050	1
16	P16	5.099	13.000	5.099	1	36	P36	17.000	24.331	17.000	1
17	P17	8.062	0.000	0.000	2	37	P37	10.198	3.606	3.606	2
18	P18	4.472	12.530	4.472	1	38	P38	5.099	12.369	5.099	1
19	P19	9.487	7.280	7.280	2	39	P39	9.055	3.606	3.606	2
20	P20	13.000	20.396	13.000	1	40	P40	13.038	20.224	13.038	1

Tabel 4. Datasets second iteration

No	Name	C1	C2	minimum distance	Clusters	No	Name	C1	C2	minimum distance	Clusters
1	P1	27.166	5.657	5.657	2	21	P21	1.000	30.017	1.000	1
2	P2	0.000	31.016	0.000	1	22	P22	11.180	20.224	11.180	1
3	P3	22.091	9.487	9.487	2	23	P23	5.099	26.000	5.099	1
4	P4	14.142	17.029	14.142	1	24	P24	20.000	11.045	11.045	2
5	P5	15.000	16.031	15.000	1	25	P25	29.275	5.385	5.385	2
6	P6	15.033	16.000	15.033	1	26	P26	7.280	24.187	7.280	1
7	P7	33.015	2.000	2.000	2	27	P27	35.228	5.000	5.000	2
8	P8	34.059	3.162	3.162	2	28	P28	6.708	25.318	6.708	1
9	P9	27.295	6.403	6.403	2	29	P29	31.016	0.000	0.000	2
10	P10	5.831	28.636	5.831	1	30	P30	25.710	9.220	9.220	2
11	P11	3.162	28.071	3.162	1	31	P31	34.015	3.606	3.606	2
12	P12	0.000	31.016	0.000	1	32	P32	22.361	10.296	10.296	2
13	P13	5.000	26.019	5.000	1	33	P33	8.944	23.537	8.944	1
14	P14	14.866	17.464	14.866	1	34	P34	25.179	6.325	6.325	2
15	P15	32.062	3.162	3.162	2	35	P35	5.099	26.077	5.099	1
16	P16	10.050	21.000	10.050	1	36	P36	2.000	33.015	2.000	1
17	P17	22.361	10.296	10.296	2	37	P37	25.080	6.708	6.708	2
18	P18	11.180	20.025	11.180	1	38	P38	10.050	21.095	10.050	1
19	P19	24.187	7.280	7.280	2	39	P39	24.021	7.280	7.280	2
20	P20	2.000	29.017	2.000	1	40	P40	2.236	29.069	2.236	1

Table 3 shows the members of cluster-1 (C1) is P2, P4, P5, P6, P10, P11, P12, P13, P14, P16, P18, P20, P21, P22, P23, P26, P28, P33, P35, P36, P38 and P40. Members of cluster-2 (C2) is P1, P3, P7, P8, P9, P15, P17, P19, P24, P26, P27, P29, P30, P31, P32, P34, P37 and P39. Total cost is total C1 + C2 = 359.541 + 396.221 = 755.762. After getting the distance from each object (cost) in the first iteration, then continue the second iteration. New medoid candidates in the second iteration can be randomly selected. In this second iteration, if the medoid formed does not change, the process is stopped. The selected medoid is P12(85, 95) as medoid-1 and P29(84, 64) as medoid-2.

C1

$$d(P1, P12) = \sqrt{(88-85)^2 + (68-95)^2} = 27.166$$

$$d(P2, P12) = \sqrt{(85-85)^2 + (95-95)^2} = 0$$

$$d(P3, P12) = \sqrt{(87-85)^2 + (73-95)^2} = 22.091$$

$$d(P5, P12) = \sqrt{(85-85)^2 + (80-95)^2} = 15$$

$$d(P38, P12) = \sqrt{(86-85)^2 + (85-95)^2} = 10.050$$

$$d(P39, P12) = \sqrt{(86-85)^2 + (71-95)^2} = 24.021$$

Submitted: **02/10/2020**; Accepted: **19/11/2020**; Published: **22/01/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 1, Januari 2021, Page 42-47
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2523

$$d(P40, P12) = \sqrt{(86-85)^2 + (93-95)^2} = 2.236$$

C2

$$d(P1, P29) = \sqrt{(88-84)^2 + (68-64)^2} = 5.657$$

$$d(P2, P29) = \sqrt{(85-84)^2 + (95-64)^2} = 31.016$$

$$d(P3, P29) = \sqrt{(87-84)^2 + (73-64)^2} = 9.487$$

$$d(P5, P29) = \sqrt{(86-84)^2 + (93-64)^2} = 16.031$$

$$d(P38, P29) = \sqrt{(86-84)^2 + (85-64)^2} = 21.095$$

$$d(P39, P29) = \sqrt{(86-84)^2 + (71-64)^2} = 7.280$$

$$d(P40, P29) = \sqrt{(86-84)^2 + (93-64)^2} = 29.069$$

Table 4 shows that the members of cluster-1 (C1) are P2, P4, P5, P6, P10, P11, P12, P13, P14, P16, P18, P20, P21, P22, P23, P26, P28, P33, P35, P36, P38, and P40. Cluster-2 members (C2) are P1, P3, P7, P8, P9, P15, P17, P19, P24, P26, P27, P29, P30, P31, P32, P34, P37, and P39. Total cost is the total C1 + C2 = 581,495 + 528,931 = 1.110,426. the member C1 in the first iteration is the same as the member C1 in the second iteration. Likewise, the C2 member in the first iteration is the same as the C2 member in the second iteration. The next step is to calculate the total deviation. The formula for deviation is the total cost (*b*) of the second iteration minus the total cost (*a*) of the first iteration. if b < a then the iteration is terminated. so that the cluster members formed in each medoid are used in the second iteration.

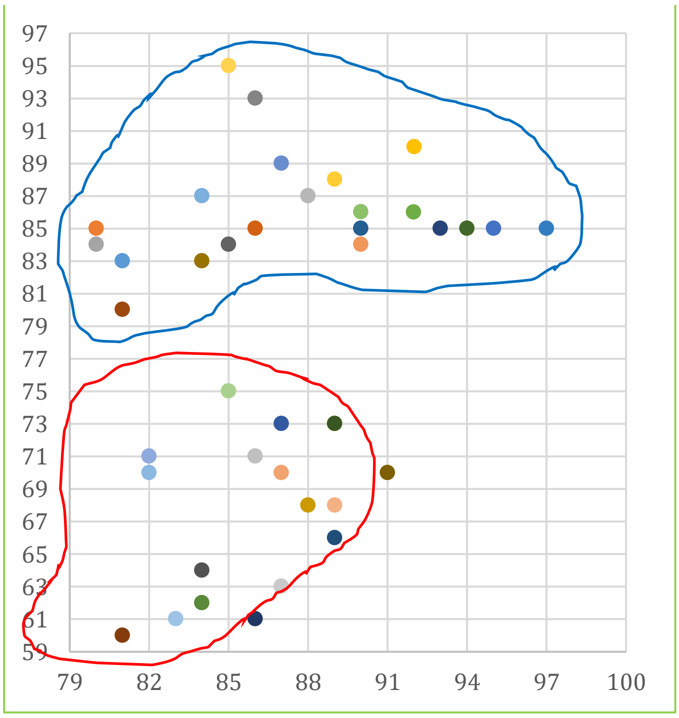

Figure 3. Distribution of clustered data

Cluster 1
Cluster 2

Figure 3 shows the data that has been clustered into two clusters. The initial data is very random successfully grouped into two groups based on initial data input.

4. CONCLUSION

The results of this study indicate that in the second iteration, the PAM method has succeeded in grouping data into 2 clusters. This happens because the cluster members do not change anymore and the value *S* is greater than zero. Dataset of 40 participants who have passed, they can be grouped into 2 clusters using the PAM method. The first cluster is formed with members P2, P4, P5, P6, P10, P11, P12, P13, P14, P16, P18, P20, P21, P22, P23, P26, P28, P33, P35, P36, P38, and P40. The second cluster (C2) are P1, P3, P7, P8, P9, P15, P17, P19, P24, P26, P27, P29, P30, P31, P32, P34, P37, and P39. The K-medoid method or the PAM method succeeded in grouping 40 students into two clusters. In Figure 3, it is clear that the difference between cluster 1 and cluster 2. Calculating the distance

Submitted: **02/10/2020**; Accepted: **19/11/2020**; Published: **22/01/2021**

Volume 5, Nomor 1, Januari 2021, Page 42-47
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2523

and iteration on the k-medoids successfully classifies the data. The cluster method is very important to implement because with the PAM method students with the same level of intelligence will be grouped into one group. This clustering process can support the teaching and learning process to be better so that the learning process can be more effective and efficient.

It is hoped that this research can be used as a reference for schools that want to divide students into several parts based on the scores obtained at school with the scores obtained at the time of selection.

REFERENCES

[1] P. G. VanderHart, "Why Do Some Schools Group by Ability?. Some Evidence from the NAEP," *Am. J. Econ. Sociol.*, vol. 65, no. 2, pp. 435–462, Apr. 2006.
[2] T. Wiedemann, "Virtual textbook for modeling and simulation," *Winter Simul. Conf. Proc.*, vol. 2, pp. 1660–1665, 2000.
[3] R. Agrawal, B. Golshan, and E. Terzi, "Grouping students in educational settings," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, 2014, pp. 1017–1026.
[4] K. Holmberg, "Formation of student groups with the help of optimisation," *J. Oper. Res. Soc.*, vol. 70, no. 9, pp. 1538–1553, 2019.
[5] S. A. A. Freitas, R. C. Silva, E. D. Canedo, and T. F. R. Lucena, "A tool for students' grouping in classroom," *Proc. - Front. Educ. Conf. FIE*, vol. 2016-Novem, no. December, 2016.
[6] R. Santos, B. N. Luz, V. F. Martins, D. R. C. Dias, M. de Paiva Guimarães, and M. D. P. Guimaraes, "Teaching-Learning Environment Tool to Promote Individualized Student Assistance," in *Conference: International Conference on Computational Science and Its Applications*, 2015, vol. 9156, no. June, pp. 143–155.
[7] J. B. Nisbet, N. J. Entwistle, B. McQuillin, and I. M. Robinson, "Staff and student perceptions of the teaching-learning environment: A case study," *Int. J. Electr. Eng. Educ.*, vol. 42, no. 1, pp. 30–40, 2005.
[8] A. Prahara, D. P. Ismi, and A. Azhari, "Parallelization of Partitioning Around Medoids (PAM) in K-Medoids Clustering on GPU," *Knowl. Eng. Data Sci.*, vol. 3, no. 1, pp. 40–49, Aug. 2020.
[9] R. D. Cahyaningrum, A. Bustamam, and T. Siswantining, "Implementation of spectral clustering with partitioning around medoids (PAM) algorithm on microarray data of carcinoma," *AIP Conf. Proc.*, vol. 1825, no. November 2019, 2017.
[10] G. Shiyamalagowri, P. Ganapathy, V. Zaalishvili, and D. Melkov, "Partitioning around medoids approach application for computation of regional flood and landslide quantiles," *E3S Web Conf.*, vol. 157, p. 02001, Mar. 2020.
[11] M. Bipul Hossen and M. Rabiul Auwul, "Comparative Study of K-Means, Partitioning Around Medoids, Agglomerative Hierarchical, and DIANA Clustering Algorithms by Using Cancer Datasets," *Biomed. Stat. Informatics*, vol. 5, no. 1, p. 20, 2020.
[12] F. R. Naibaho, "Optimalisasi Keputusan Untuk Memperoleh Keuntungan Maksimal Dalam Penentuan Kelayakan Operasional Kendaraan Dengan Menggunakan Metode Dynamic Programming Pada PT. Tor Ganda," 2015.
[13] Y. Zhao, X. Liu, and H. Zhang, "The K-Medoids Clustering Algorithm with Membrane Computing," *TELKOMNIKA Indones. J. Electr. Eng.*, vol. 11, no. 4, pp. 2050–2057, 2013.
[14] F. R. Naibaho, "FUZZY LOGIC METODE TSUKAMOTO UNTUK PREDIKSI PRODUKSI CPO DENGAN PERMINTAAN BERSIFAT STOKASTIK PADA PT. TOR GANDA," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, Nov. 2019.
[15] A. Zandi and E. Roghanian, "Extension of Fuzzy ELECTRE based on VIKOR method," *Comput. Ind. Eng.*, vol. 66, no. 2, pp. 258–263, Oct. 2013.

Submitted: **02/10/2020**; Accepted: **19/11/2020**; Published: **22/01/2021**