CLUSTERING LARGE APPLICATION USING METAHEURISTICS (CLAM) FOR GROUPING DISTRICTS BASED ON PRIMARY SCHOOL DATA ON THE ISLAND OF SUMATRA

Naura Ghina As-shofa[1], Vemmie Nastiti Lestari[2*]

[1,2]*Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, Indonesia*
Email: [1]nauraghina99@mail.ugm.ac.id, [2]vemmie.lestari@ugm.ac.id
[*]Corresponding Author

Abstract. K-medoids is one of the partitioning methods with the medoid as its center cluster, where medoid is the most centrally located object in a cluster, which is robust to outliers. The k-medoids algorithm used in this study is Clustering Large Application Using Metaheuristics (CLAM), where CLAM is a development of the Clustering Large Application based on Randomized Search (CLARANS) algorithm in improving the quality of cluster analysis by using hybrid metaheuristics between Tabu Search (TS) and Variable Neighborhood Search (VNS). In the case study, the best cluster analysis method for classifying sub-districts on the island of Sumatra based on elementary school availability and elementary school process standards is the CLAM method with k=6, num local = 2, max neighbor = 154, tls = 50 and set radius = 100-10:5. It can be seen that based on the overall average silhouette width value, the CLAM method is better than the CLARANS method.

Keywords: CLAM, CLARANS, Tabu Search, Variable Neighborhood Search.

I. INTRODUCTION

Cluster analysis is the process of grouping objects into several groups or clusters so that objects in a cluster have a high level of similarity (homogeneity) and objects between clusters have a high level of difference (heterogeneity). Differences and similarities are assessed based on specific characteristics or variables that describe things and are measured using distance measures [1].

One of the most popular methods is non-hierarchical cluster analysis (non-hierarchical methods) or partitioning. In this method, the number of clusters is determined from the start by the researcher, and each group has a cluster center point (centroid). The most popular and frequently used partitioning methods are the k-means and k-medoids. The k-means method uses the average data value in each cluster formed, while the k-medoids method uses the medoid or object most centrally located in a cluster. Because the average is a central measure that is not robust to outliers, using the average as the cluster center in the k-means method makes the method more sensitive to the presence of outliers than the k-medoids method [2].

Various algorithms can be used in the k-medoids method, including Partitioning Around Medoids (PAM), Clustering Large Applications (CLARA), and CLARANS [3]. This research uses the CLAM algorithm to develop the CLARANS algorithm. CLAM applies a hybrid

metaheuristic between TS and VNS to improve the quality of the CLARANS method in cluster application analysis on large datasets containing outliers [3]. In the CLAM method, a tabu list parameter is used, which stores search history to prevent wasted search efforts by revisiting previously visited neighbors. This parameter overcomes the shortcomings of the CLARANS method, which has an interconnected search space so that many neighbors can be seen several times. In the CLAM method, a linear reduction schedule is also used, which determines the radius of the medoid at each stage. At the beginning of the algorithm, suppose the node is far from the resting position, namely the set of k objects, which is the best choice for the medoid. Therefore, at the start of the CLAM algorithm, each medoid can swap with anything in the data set. However, in the next stage of the algorithm, the k-medoids will move closer to the resting position so that each medoid is only allowed to exchange with data objects that are within a reduced radius until, at the end of the algorithm, the medoids are only allowed to trade with their nearest neighbors. This is done because if the k-medoids are within a certain distance from the resting position, large movements or exchanges of medoids with objects far from the medoid position are no longer efficient. Reducing the radius at certain stages in the CLAM algorithm improves the CLARANS algorithm, which does not limit the range of movement at any stage.

The grouping that will be carried out is based on school data, especially elementary schools. Elementary school is an educational institution that is formally organized and lasts six years to complete 6 grade levels, from grade 1 to grade 6. The maximum number of students in one study group is determined based on Minister of Education and Culture Regulation Number 22 of 2016 concerning Primary and Secondary Education Process Standards. This policy aims to create a comfortable learning atmosphere to make the learning process effective [4]. The grouping is carried out to determine which regions have and have not given rights to citizens of school age at the elementary school level to receive a decent education. In this research, grouping was carried out using CLAM in sub-districts on the island of Sumatra. The author assesses the feasibility of the education provided based on educational process standards, namely the student-teacher ratio (RSG) and study group student ratio (RSR), in addition to the number of schools in a sub-district will be considered.

II. METHODOLOGY

2.1. Metaheuristik

The term metaheuristics uses a combination of two Greek words: meta, which means high-level or advanced methodology, and heuristic, which means the art of finding strategies to solve problems [5]. Metaheuristics is a computational approach to finding optimum or near-optimum solutions to optimization problems by trying iteratively to improve candidate solutions by considering the desired solution quality limits [6]. The metaheuristic approach has general characteristics, including having the ability to solve hard combinatorial problems with problems of relatively large size and competitive computing time, not using gradient calculations from the objective function. Most metaheuristic techniques generate several candidate solutions in each iteration (population-based, such as Genetic Algorithm, Particle Swarm Optimization, Ant Colony). However, there is a metaheuristic technique that only generates one solution in each iteration, namely Simulated Annealing. Tabu Search (TS) and Variable Neighborhood Search (VNS) are the most popular metaheuristics. This algorithm is the algorithm that is the basis for grouping in CLAM. CLAM applies a hybrid metaheuristic between TS and VNS.

2.2. *Tabu Search* (TS)

Tabu search is a metaheuristic that guides local heuristic search procedures to explore the solution space outside the local optimum [7]. Tabu search can overcome one of the shortcomings of simple local search: it does not record any search history, so that some search attempts can be repeated. In the tabu search method, there are several terms used, namely:
1. Tabu list contains solutions that have been visited (tabu-active).
2. Aspiration criteria are specific criteria or conditions that allow a taboo movement to be carried out.
3. Intensification (medium-term memory) is a medium-term memory that stores several quality solutions (elite solutions) produced during the search process.
4. Diversification (long-term memory) is a long-term memory that stores information about candidate solutions that have been visited.
5. Tabu tenure or taboo list size is the duration for which a movement remains tabu.

The algorithm or steps used in tabu search are as follows [8].
1. Initialization *tabu list* $= \emptyset$.
2. Choose a solution x as an initial solution drawn randomly from the data set.
3. Determining several candidate solutions is done by moving from the initial solution x and creating a list of candidate solutions.
4. Calculate all objective function values of the candidate solutions and determine the best solution x' with the most optimal accurate function value among all candidate solutions (minimized or maximized).
5. Check whether the best solution x' is contained in the tabu list. If x' is in the tab list, check the aspiration criteria. Meanwhile, if the opposite is true, determine $x = x'$, update the tabu list, and continue to step 7.
6. Check whether the movement meets the aspiration criteria. If the move meets the aspiration criteria, then $x = x'$, update the tabu list, and go to step 7. Meanwhile, if vice versa, delete x' and return to step 4.
7. Check whether the termination criteria have been met. If the stopping criteria are met, the output is x. If otherwise, go back to step 3.

2.3. *Variable Neighborhood Search* (VNS)

VNS is a metaheuristic proposed by Mladenovic and Hansen in 1997. VNS is based on using more than one environmental structure and changing this structure systematically during local search [9].

VNS consists of three stages: shaking, improvement, and neighborhood change. The environmental structure is selected before starting these three stages: the number of environmental structures (starting now denoted by k_{max}), each object or node, and the stopping criteria. Then, these three stages are carried out alternately until the termination criteria are met. Two different notations differentiate the environments used in the shaking and improvement stages, namely $\aleph$ and N . For example, $\aleph = \{\aleph_1, \ldots, \aleph_{kmax}\}$ with $1 \leq k \leq k_{max}$, then $\aleph$ is defined as the set of k_{max} environmental structures, and $\aleph_k(x)$ is defined as the kth environmental structure where in that environment there is a set of solutions x while the notation The environment used in the improvement stage is $N(x)$ which is the neighbor of

solution x, where neighbors of solution x are defined as objects or nodes that are in x's environment.

2.4. *Clustering Large Application Using Metaheuristics* (CLAM)

CLAM is a k-medoids cluster analysis method that applies a hybrid metaheuristic between Tabu Search (TS) and Variable Neighborhood Search (VNS). Hybrid metaheuristics are used to improve the quality of the CLARANS method in cluster analysis applications with large amounts of data. The limit used to be able to categorize large amounts of data is 1000 data [10].

The CLAM algorithm is not much different from the CLARANS algorithm in that both use max neighbor constraints and iteration constraints (num local), only in the CLAM algorithm there is the use of a tabu list which stores the search history and a linear reduction schedule which determines the radius of the medoid at each stage.

The steps for the Clustering Large Application Using Metaheuristics (CLAM) algorithm are as follows:

1. Determine k as the number of clusters you want to form.
2. Calculates diameter as the farthest distance between two objects.
3. Define set radius, num local, maxneighbor, and tabu list size (tls).
4. Initialize the objective function of the best solution ($f(bs)$) with large numbers, tabu list = Ø and tabu list length (TLL) = 0
5. Initialize the number of checks for a node with the value one or $i = 1$.
6. Select incumbents randomly from the data set to get medoids representing several clusters.
7. Add the incumbent to the tabu list and add a value of 1 to the TLL.
8. Initialize the radius of the medoid in step 1 with the value one or radius_index = 1.
9. Initialize the number of neighbors of a node being checked with the value one or $j = 1$.
10. Randomly select neighbors of incumbents not on the tabu list.
11. If TLL > TLS, the first node in the tabu list is deleted, and the neighbor is added to the tabu list.
12. Calculate the objective function values of the incumbent and neighbors ($f(in)$ and $f(nb)$).
13. Compare the values of $f(nb)$ and $f(in)$. If $f(nb) < f(in)$, swap the incumbent with a neighbor and return to step 9. If $f(nb) \geq f(in)$, add the value 1 to j and return to step 10.
14. If j > maxneighbor, add a value of 1 to radius_index and return to step 9.
15. If radius_index > t, compare $f(in)$ with $f(bs)$. If $f(in) < f(bs)$, swap the best solution with the incumbent.
16. Adds the value 1 to i. If i > numlocal, it returns the best solution, and the process stops. If not, return to step 6. The CLAM algorithm then repeats to search for another numlocal until it is satisfied.

III. CASE STUDY

The data used in this case study is secondary data regarding elementary school (SD) data for the 2021/2022 even semester academic year in 1949 sub-districts on Sumatra Island. Data will be grouped based on three characteristics or variables, namely student-teacher ratio (RSG), study group-student ratio (RSR), and number of schools [11]. Data sourced from the Basic Education Data website of the Ministry of Education, Culture, Research and Technology

(dapo.kemdikbud.go.id) [12]. Before testing assumptions or analysis, standardization is carried out first on the data. The assumption test that needs to be carried out is that the data is representative and there is no multicollinearity between variables. The data is considered representative because the data can describe the actual conditions of the existing sub-districts. Test the assumption that there is no multicollinearity between variables. The test is carried out by looking at the Variance Inflation Factors (VIF) value for each variable, and it is said that there is no multicollinearity if the VIF < 10 for each variable [13]. The VIF value can be calculated using the following formula.

$$VIF_k = \frac{1}{(1 - R_k^2)} \tag{1}$$

where VIF_k is the VIF value of the kth variable, and R_k^2 is the coefficient of determination obtained if the kth variable is regressed against other independent variables. Based on tests using equation (1), the VIF results obtained are presented in Table 1.

Table 1. VIF value for each Variable

Feature	VIF
Z_{RSG}	5.721144
Z_{RSR}	5.384213
$Z_{Sekolah}$	3.028149

Based on Table 1, it can be seen that each variable has a VIF value < 10, so it can be concluded that the assumption of no multicollinearity between variables is met.

Because the assumptions have been met, the elbow criterion is used to determine the optimal number of clusters. The elbow criterion can be used by looking at the plot between the SSE value and the number of clusters formed. If a drastic decrease is seen and an elbow is formed for the SSE value at a k value, then the k value is the number of clusters that will be formed in the cluster analysis. The following is a plot between the SSE value and the number of k clusters formed.
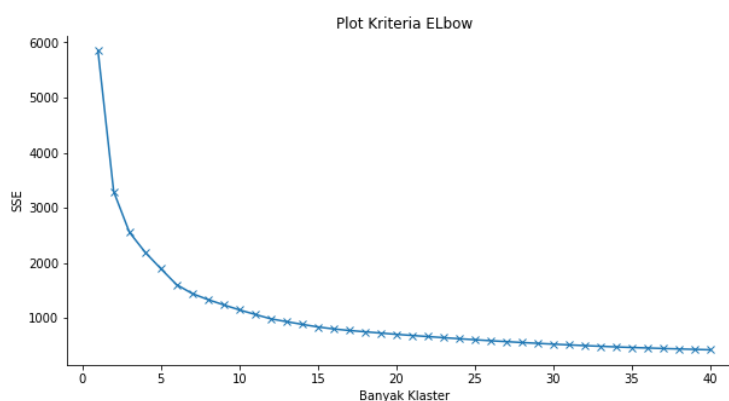

Fig. 1. Elbow Criterion Plot

Based on the plot in Figure 1, it can be seen that there is a decrease in the SSE value as the number of clusters formed increases. It can be seen that when $k = 6$, the plot forms an elbow, and then for k values, the plot shows a relatively stable SSE value. According to researchers, based on this plot, the number of clusters that will be formed is 6.

The objects were grouped into 6 clusters ($k = 6$) using the CLAM method. The distance measure that will be used is Euclidean distance. In the CLAM simulation, four parameters are used, namely num local of 2, a max neighbor of 145 to 174 (1.25% to 1.5% of $k(n − k)$) and two other parameters, namely tls and set radius, which will be determined based on the results of the experiments to be carried out. CLAM simulation was carried out three times for each parameter combination for each experiment. The grouping results with the highest validation values were selected from the three simulations. The results of all maxneighbors will be compared, and the maxneighbor with the best cluster analysis results will be selected. Four experiments were carried out with tls parameter values, and the set radius can be seen in Table 2.

Tabel 2. Experiments based on tls and setradius Parameter Values

Eksperiment	CLAM	
	tls	*setradius*
1	50	100-10:5
2	100	100-10:5
3	50	50-10:5
4	100	50-10:5

In the first experiment, the parameter $tls = 50$ means that the tabu list size is 50 or that each neighbor will not be revisited in 50 iterations. Meanwhile, the parameter $setradius = 100$-10:5 can be interpreted as using start_radius = 100% of dim, end_radius = 10% of dim, and t = 5 steps, then the settings that will be used in the linear reduction schedule are (100%, 77.5%, 55%, 32.5%, 10%) of dim.

3.1 Comparison of CLAM using 4 Experiments

Based on the simulations carried out on the four experiments, the best overall average silhouette width value was obtained for the four experiments, namely first at maxneighbor = 154, namely 0.294, second at maxneighbor = 148, namely 0.277, third at maxneighbor = 149, namely 0.278 and fourth at maxneighbor = 146, which is 0.28. The experiment that produces the highest overall average silhouette width value or closest to 1 is the best. A comparison graph of the overall average silhouette width values in the four experiments can be seen in Figure 2.
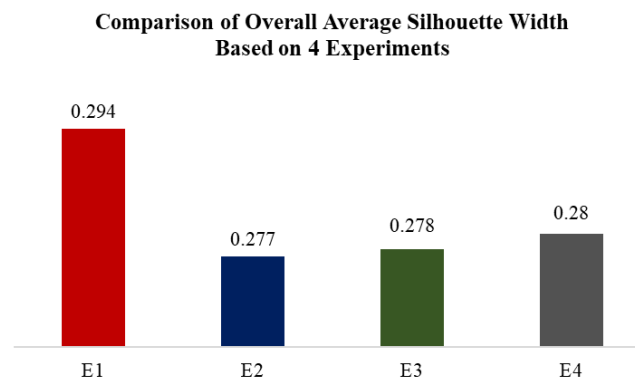

Fig. 2. Comparison graph of overall average silhouette values
width CLAM with four experiments

Based on Figure 2, it is known that grouping the CLAM method into 6 clusters for experiment 1 has the highest overall average silhouette width value of 0.294 compared to other experiments, so the CLAM method will be used with a TLS of 50 and a radius of 100-10:5 for analyzing case studies in this research.

3.2 Profiling Cluster Results

Comparison of characteristics between clusters can be determined by comparing the total scores between medoids for student-teacher ratio (RSG) and group-student ratio (RSR) and paying attention to the number of schools in each medoid. Sequentially, a score of 1 to 6 is given to each cluster with the lowest to highest RSG and RSR values. The following table compares the scores for each cluster for the RSG and RSR values.

Table 3. Comparison of Total Scores between Clusters for RSG and RSR

Medoid	Variable Value		Variable Score		Score Total
	RSG	RSR	RSG	RSR	
1	14.5667	19.8636	4	4	8
2	13.0753	18.7077	3	3	6
3	9.94792	12.9054	1	1	2
4	18.6194	22.6818	6	5	11
5	11.3156	13.9126	2	2	4
6	17.5077	24.2447	5	6	11

Based on Table 3, it is known that the highest scores are clusters 4 and 6, with a total score of 11, so that clusters 4 and 6 are the clusters that have the highest RSG and RSR. Furthermore, it can be seen in Table 4, the number of sub-districts for each cluster by Province.

Table 4. Number of Districts in each Cluster based on Province

Province	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Total District
Aceh	13	116	107	179	8	2	289
North Sumatra	81	100	76	73	29	61	450
West Sumatra	45	35	25	45	28	15	179
Riau	26	32	8	32	15	26	169
Jambi	23	37	30	43	6	10	142
South Sumatra	47	52	31	23	4	26	241
Bengkulu	7	64	37	44	2	0	128
Lampung	51	53	9	19	13	21	228
Bangka Belitung Islands	2	7	2	2	1	6	47
Riau Islands	0	14	39	33	2	6	76

Minister of Education and Culture Regulation Number 23 of 2013 article 2 paragraph (2) states that each SD/MI has 1 (one) teacher for every 32 students. The maximum permitted student-teacher ratio (RSG) is 32. Minister of Education and Culture Regulation Number 22 of

2016 concerning Process Standards states that the maximum number of students in one study group for elementary school education is 28 students [4]. This means the maximum permitted study group-student ratio (RSR) is 28. Cluster 4 and Cluster 6 are the clusters that have the highest RSG and RSR values. Therefore, the suitability of the RSG and RSR values will be checked on the two clusters. From the two clusters, it is known that all sub-districts have RSGs smaller than 32, so all sub-districts have fulfilled the regulations regarding RSG.

Meanwhile, for the RSR value, it is known that of the total of 37 sub-districts that have an RSR value above 28 (maximum RSR value), there are 29 sub-districts included in cluster 4 and 8 other sub-districts included in cluster 6. This can be interpreted as there are still sub-districts that need to comply with government regulations regarding RSR and receive more attention so that students in the sub-district can receive a more effective learning process. It is necessary to know the cause of the high RSR in the sub-district, whether too many students want to go to school in the sub-district so that the number of students is enormous or there needs to be more class units available.

3.3 Comparison of CLAM and CLARANS

It is known that the CLAM algorithm is a development of the CLARANS algorithm. Therefore, in this case study, we will compare cluster analysis results using the CLAM and CLARANS algorithms. Comparisons are made involving all data and data with a proportion of outliers using the Euclidean distance measure. For data with outliers, experiments were carried out on data that did not contain outliers and data that contained outliers in proportions of 1%, 2%, 3%, and 4%. The respective amounts of outlier and non-outlier data used are presented in Table 5.

Table 5. Proportion of Outliers

Proportion of Outliers	Number of Outliers	Data Size
0%	0	1843
1%	19	1843
2%	38	1843
3%	57	1843
4%	77	1843

In this comparison, grouping simulations were carried out three times using the CLARANS and CLAM algorithms. In the CLARANS algorithm, data is grouped into 6 clusters with the numlocal and maxneighbor parameter values used by the best parameter values in the CLAM algorithm, namely numlocal of 2 and maxneighbor of 154. In the CLAM algorithm, tls = 50 and set radius = 100-10: 5. The best grouping results for the two algorithms are compared by looking at each algorithm's best overall average silhouette width value. Below is a comparison table of the best overall average silhouette width values from the two algorithms for all data and data with a proportion of outliers.

Table 6. Comparison of Overall Average Silhouette Width Values on CLAM and CLARANS

	CLAM	CLARANS	Proportion of Outliers	CLAM	CLARANS
All Data	**0.294**	0.235	0%	**0.2753705**	0.265977
			1%	**0.3020656**	0.257695
			2%	**0.2741035**	0.254549
			3%	**0.2878553**	0.285342
			4%	**0.2888027**	0.265659

Based on Table 6, it is known that from the three simulations that have been carried out on both algorithms with a total of 6 clusters for all data and data with a proportion of outliers, it can be seen that the CLAM method obtains an overall average silhouette width value that is closest to 1 compared to the CLARANS method. Therefore, it can be concluded that in this case study, the CLAM method is more effective than the CLARANS method by using all data when the data does not contain outliers (0%) or when the data contains outliers with a proportion of 1% - 4%.

IV. CONCLUSION

Based on the results of the case study analysis that has been carried out, it is concluded that from the four experiments using the CLAM method, it is known that the CLAM method with several clusters of 6, num local of 2, a max neighbor of 154, tls of 50 and set radius = 100-10:5 is the parameter choice best in applying the CLAM method to group sub-districts on the island of Sumatra based on information about Elementary Schools (SD). Furthermore, cluster 4 and cluster 6 are the clusters that have the highest RSG and RSR values. From the two clusters, it is known that all sub-districts have RSGs smaller than 32, so all sub-districts have fulfilled the regulations regarding RSG. Meanwhile, for the RSR value, it is known that of the total of 37 sub-districts that have an RSR value above 28 (maximum RSR value), there are 29 sub-districts included in cluster 4 and 8 other sub-districts included in cluster 6. This can be interpreted as there are still sub-districts that have not complied with government regulations regarding RSR and need to receive more attention so that students in the sub-district can receive a more effective learning process. From the three simulations that have been carried out with a total of 6 clusters for all data and data with a proportion of outliers, it can be seen that the CLAM method obtains an overall average silhouette width value closest to 1 compared to the CLARANS method. Therefore, it can be concluded that in this case study, the CLAM method is more effective than the CLARANS method by using all data when the data does not contain outliers (0%) or when the data contains outliers with a proportion of 1% - 4%.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, USA: Elsevier Inc, 2012.

[2] Y. Flowrensia, "Perbandingan Penggerombolan K-Means dan K-Medoid pada Data yang Mengandung Pencilan," FMIPA IPB, Bogor, 2010.

[3] Q. H. Nguyen and V. J. Rayward-Smith, "CLAM: Clustering Large Applications Using Metaheuristics," *J. Math Model Algorithm,* vol. 10, pp. 57-78, 2011.

[4] "Peraturan Menteri Pendidikan dan Kebudayaan Nomor 22 Tahun 2016 tentang Standar Proses pendidikan Dasar dan Menengah," Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi, 2016. [Online].

[5] F. Glover, "Future Paths for Integer Programming and Links to Artificial Intelligence," *Computer and Operations Research,* vol. 13, pp. 533-549, 1986.

[6] B. Santosa and T. J. Ai, Pengantar Metaheuristik Implementasi dengan Matlab, Surabaya: ITS Tekno Sains, 2017.

[7] F. Glover, M. Laguna and R. Marti, "Principle of Tabu Search," in *Handbook of Approximation Algorithms and Metaheuristics*, 2007.

[8] F. Glover, M. Laguna and R. Marti, "Tabu Search: A Tutorial," *Interfaces,* vol. 20, no. 4, pp. 74-94, 2009.

[9] P. Hansen, N. Mladenovic, R. Tojosijevic and S. Hanafi, "Variable Neighborhood Search," *EURO Journal in Computational Optimization,* vol. 5, pp. 3-35, 2016.

[10] T. Ng Raymond and J. Han, "A Method for Clustering Objects for Spatial Data Mining," *IEEE Transactions on Knowledge and Data Engineering,* vol. 14, no. 5, pp. 1003-1016, 2002.

[11] N. S. Perdana, M. Handayani and J. Purnama, "Analisis Hubungan Rombongan Belajar dan Siswa Per Rombel dengan Mutu Lulusan," Jakarta, Pusat Penelitian Kebijakan Badna Penelitian, Pengembandan dan Perbukuan Kementerian Pendidikan dan Kebudayaan, 2020.

[12] "Data Pokok Pendidikan," Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi, [Online]. Available: dapo.kemdikbud.go.id. [Accessed 17 May 2022].

[13] J. F. J. Hair, R. E. Anderson, B. J. Babin and W. C. Black, Multivariate Data Analysis, New York: Macmillan: Perason Education Limited, 2014.

[14] K. S. Al-sultan, "A Tabu Search Approach To The Clustering Problem," *Pattern Recognition,* vol. 28, no. 9, pp. 143-1451, 1995.

[15] L. J. Bain and M. Engelhardt, Introduction to Probability and Mathematical Statistics, California: Duxbury Press, 1992.

[16] G. Brock, V. Pihur and S. Datta, "clVAlid: An R Package for Cluster Validation," *Kournal of Statistical Software,* vol. 25, no. 4, pp. 1-22, 2008.

[17] L. Kaufman and P. J. Rousseeuw, Finding Groups in Data : An Introduction to Cluster Analysis, New Jersey: John Wiley and Sons, Inc, 1990.

[18] M. H. Kutner, C. J. Nachtsheim and W. Li, Applied Linear Statistical Models, New York: Mc Graw-Hill, 2005.

[19] N. Mladenovic and P. Hansen, "Variable Neighborhood Search," *Computers & Operations Research,* vol. 24, no. 11, pp. 1097-1100, 1997.

[20] A. B. Muslim, "Analisis Klaster Menggunakan Metode CLARA pada Data yang Mengandung Pencilan," FMIPA UGM, Yogyakarta, 2018.

[21] [Online]. Available: http://searchcode.com/codesearch/view/3007170/. [Accessed 24 April 2022].

[22] E. G. Talbi, Metaheuristic: From Design to Implementation, United States of America: John Wiley & Sons, Inc, 2009.

[23] A. V. Gunawan, "Metode K-Medoids dengan Algoritma CLARANS pada Dataset Besar dengan Pencilan," FMIPA UGM, Yogyakarta, 2018.

[24] K. D. Yulianingsih, "Analisis Klaster Menggunakan Algoritma I-CLARANS untuk Dataset Besar dengan Pencilan," FMIPA UGM, Yogyakarta, 2019.