JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 12 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v12i3.4862

TOWARDS OPTIMIZATION: A DATA-DRIVEN APPROACH USING K-MEDOIDS CLUSTERING ALGORITHM FOR REGIONAL EDUCATION QUALITY ASSESSMENT

Harun Al Azies[1,2*], Fawwaz Atha Rohmatullah[3], Hani Brilianti Rochmanto[4], Devi Putri Isnarwaty[5]

[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Semarang, Indonesia
[2]Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Semarang, Indonesia
[3]Study Program in Information Systems, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Semarang, Indonesia
[4]Department of Statistics, Faculty of Science and Technology, Universitas PGRI Adi Buana, Surabaya 60234, Indonesia
[5]Smart Eco Campus Development Unit, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

Received: 11 Juli 2024
Accepted: 31 Juli 2024
Published: 7 Agustus 2024

Keywords:
Education Quality;
K-Medoids Clustering;
Machine Learning.

Corespondent Email:
harun.alazies@dsn.dinus.ac.id

Abstract. *This study applies the k-medoids clustering machine learning approach to assess regional clustering in Indonesia based on educational quality. Data on the quality of education, including indicators of school enrollment rate (APS), gross enrollment rate (APK), and pure participation rate (APM), is gathered and processed from all provinces in Indonesia. The k-medoids clustering technique is used to carry out the clustering process, while metrics like Dunn's index, connection coefficient, and silhouette score are used to evaluate the results. The study's findings indicate that three clusters are the ideal amount, with a silhouette score of 0.2388, a connectivity coefficient of 7.1405, and a Dunn's index value of 0.1651. Cluster homogeneity is likewise moderate, despite the regions' moderate distances from one another. This assessment offers a thorough understanding of Indonesia's educational quality clustering pattern, which can serve as a foundation for developing education strategies in different areas.*

1. INTRODUCTION

Education is crucial to a country's development since it is the fundamental pillar in the development of quality human resources [1]. The examination of educational quality is a vital component in determining a region's educational system's effectiveness and efficiency [2], [3]. This study focuses on three key variables to better understand and analyze educational quality: school enrollment rate, gross enrollment rate, and pure participation rate. The APS shows how many people in a given area are enrolled in school [4], the proportion of school-age children in an age group that corresponds to a particular level of education is known as the gross enrollment rate, or APK [5]. The proportion of students in a certain group who attend school at a level acceptable for their age group is known as the pure participation rate, or APM (Hasan, 2022). By considering these aspects, this study aims to provide a more comprehensive understanding of the quality of education in diverse places. This comprehension includes not only quantitative factors, but also qualitative ones such as participation, achievement, and dropout difficulties at various levels of education [7]. It is hoped that by looking further into these characteristics, this research would provide a more holistic and contextual understanding of educational quality, supporting the creation of education policies that are more sensitive to the needs of local communities [8].

This study suggests employing machine learning approaches, specifically the k-medoids clustering algorithm, to cluster regions based on education quality criteria to better understand this difficulty[9]. This method is an example of unsupervised learning [10], which allows researchers to find patterns and relationships in data without the need for manual guidance [11]. This conclusion is based on machine learning's ability to handle complex data, provide analytical efficiency, and produce objective results. This approach is bolstered by the use of education quality indicators such as APS, APK, and APM, which provide the quantitative dimensions required for a comprehensive understanding of participation, achievement, and dropout problems at various levels of education. Several examples of previous studies serve as the foundation and inspiration for the use of machine learning, namely k-medoid clustering, in analyzing educational quality. The research of [12] on the application of the K-means clustering method to the distribution of high school teachers demonstrates the success of a similar strategy in the context of educational resource allocation. The findings provide a clearer picture of the distribution of high school teachers across Indonesia, inspiring a similar method to understanding the features of education quality at the regional level. Furthermore, a study by [13] that uses K-means cluster analysis to categorize provinces in Indonesia based on community welfare indicators provides insight into the possibilities of cluster analysis in the context of development strategy. In this dimension, grouping regions based on education quality can provide a more in-depth knowledge of the differences in characteristics and educational needs among regional clusters. [14] research, which uses the K-Means Clustering method to categorize provinces in Indonesia based on education variables, demonstrates that cluster analysis can provide a deeper knowledge of the dynamics of education. In the context of this study, machine learning approaches, particularly k-medoid clustering, can help to provide a larger and deeper understanding of educational quality in diverse regions.

As a result, this study represents a significant advancement in the development of educational quality analysis methods. The application of the k-medoids clustering method, a machine learning strategy that is not yet widely employed in the context of analyzing the quality of education at the regional level, distinguishes it [15]. This research intends to provide a more in-depth and holistic picture of the dynamics of education quality in Indonesia by integrating the sophistication of this technology. By applying clustering results to the setting of Indonesian provinces, this study intends to not only discover regional groups but also obtain a better knowledge of the characteristics of each cluster. Furthermore, the project's purpose is to create regional visualizations that stakeholders can easily understand so that they can serve as a solid platform for decision-making when designing more adaptive education policies. The ramifications of this research are far-reaching for local governments, educational institutions, and society as a whole. Regional clustering and visualization findings can provide a clearer picture of the differences and similarities in education quality indicators across provinces, allowing stakeholders to identify areas that require more widespread attention. Thus, it is believed that this research would make a tangible contribution to the development of more effective education policies and have a favorable impact on enhancing overall educational quality.

2. MATERIALS AND METHODS

The study framework utilized to identify and analyze regional clustering based on education quality in Indonesia will be detailed in the Materials and Methods part of this research. Figure 1 depicts the entire process, which includes four major stages: data collecting, data preprocessing, clustering, reviewing cluster findings, and profiling. Each phase in this research methodology is intended to provide in-depth knowledge of educational situations in diverse regions and to serve as a foundation for more focused adjustments in education policy.

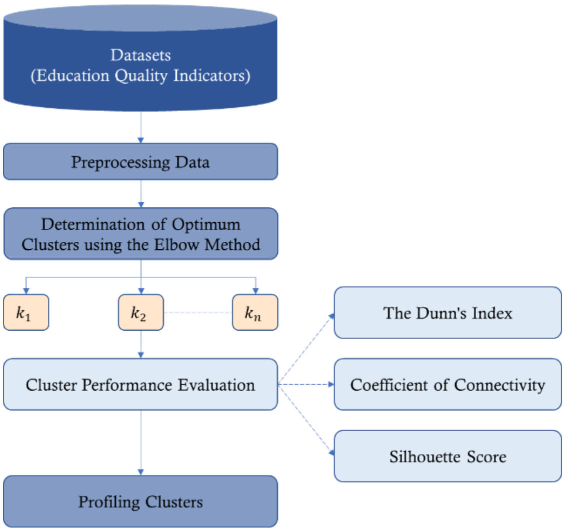

Figure 1. Research Framework for Education Quality Clustering in Indonesia

2.1. Data collection

Data on the quality of education at the elementary (SD), junior high school (SMP), and senior high school (SMA) levels from each province in Indonesia was collected for clustering analysis. Data collection includes main indicators such as APS, APK, and APM for each level of education [14]. The data source used in this research comes from the Statistics Indonesia (BPS) for 2023. This data provides a holistic picture of the condition of education in various regions of Indonesia during that period.

2.2. Data Pre-Processing

Following data collection, pre-processing is performed to ensure the data's quality and eligibility for clustering analysis. This stage involves data normalization, which tries to show the data on a uniform scale, ensuring that the variables contribute to the clustering process in a balanced manner [16]. The purpose is to prepare the data so that it may be clustered more effectively and accurately.

2.3. Clustering approach

The clustering approach is k-medoid clustering, which is a machine-learning method [17], [18]. This algorithm was chosen for its ability to deal with data containing outliers [9]. Based on education quality measurements, the clustering approach is utilized to find regional group patterns with comparable characteristics. The first stage involves determining the optimum number of clusters using the elbow strategy to ensure optimal clustering results [19]. To identify the best number of clusters (k) using the elbow technique, first determine the range of k values you want to test, ranging from 2 to the maximum number of clusters that suit the data. Then, for each k, compute the inertia value (I_k), which is the sum of the squares of the distance between each data point and its cluster center. The inertia formula is as follows[20]:

$$I_k = \sum_{i=1}^{n} min_{j=1}^{k} \left\| x_i - c_j \right\|^2 \qquad (1)$$

n is the number of data points and k is the number of clusters being tested, x_i is the *i-th* data point, c_j is the center of the cluster and $\|.\|^2$denotes the square of the Euclidean distance. The inertia value is then plotted against the number of clusters, and when adding clusters no longer results in a substantial drop in inertia, the graph forms an elbow, which is referred to as an elbow. The k value is determined as the best number of clusters for clustering analysis at this elbow point [21] .

2.4. Performance Evaluation

In this study, the cluster evaluation process was carried out to establish the appropriate number of clusters and quantify clustering quality. To evaluate cluster results, numerous metrics are utilized, including Dunn's Index, Connectivity Coefficient, and Silhouette Score [22].

a. **Dunn's Index:** Dunn's index is a tool for determining the degree of separation between clusters. A high Dunn's index value suggests higher group separation. The Dunn's Index value is utilized as a reference in this study to measure the extent to which clustering can form diverse groups well [23], [24].
b. **Coefficient of Connectivity:** The homogeneity of a cluster is assessed using the connection coefficient. A lower Connectivity Coefficient value suggests that the educational quality within the cluster is very similar. The Connectivity Coefficient is employed in this study to guarantee that regional groupings have comparable educational attributes [25].
c. **Silhouette Score:** The degree to which an observation resembles the cluster in which

it is positioned is gauged by the silhouette score [26], [27]. A favorable match level is indicated by a higher Silhouette Score rating. The Silhouette Score is employed in this study to evaluate the degree to which regions are clustered [28].

2.5. *Profiling*

Upon obtaining the best clustering outcomes, the last stage is to profile every cluster. To create a profile, each cluster's distinctive qualities are examined using measures of the level of schooling. A thorough analysis of elementary, middle, and high school APS, APK, and APM data was conducted to identify the distinctions and commonalities among the clusters. With this profiling, we hope to shed more light on the educational circumstances unique to each regional group. The findings can be utilized to inform decision-making in the pursuit of raising the standard of education throughout Indonesia, as well as to create more focused educational policies and comprehend educational development patterns in different regions.

3. RESULTS AND DISCUSSION

A critical stage in clustering analysis is determining the optimal number of clusters. The elbow method is a popular method for determining the optimal number of clusters. The Elbow approach is being used in this study to determine the point at which adding clusters does not provide a substantial boost in the explanation of data variability [19]. The investigation starts with running the k-medoids clustering method with different cluster counts and then measuring the overall dispersion or inertia value. This inertia reflects the distance between each point in the cluster and its medoid. Because each cluster is more compact, inertia tends to decrease as the number of clusters grows. The elbow point, on the other hand, is the point at which the inertia reduction begins to stall dramatically, indicating that adding more clusters delivers less advantage.

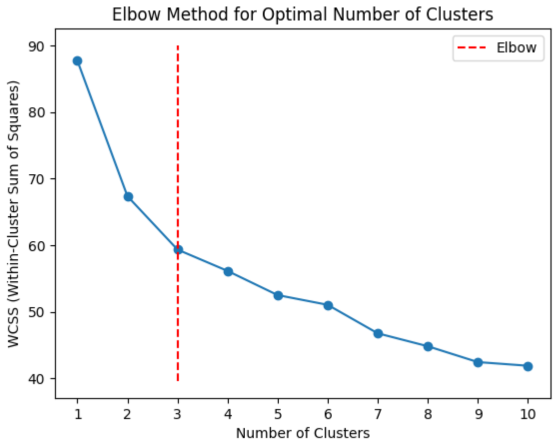

Figure 2. Elbow Method for Determining the Optimal Number of Clusters in Educational Quality Analysis

The elbow point is found at cluster number 3 according to the results of the elbow method analysis using the the Within-Cluster Sum of Squares (WCSS) shown in Figure 2 [29]. The optimal image proves this, as it shows a significant drop in inertia at that point [30]. The WCSS score assesses how effectively the clusters differ from one another. As the number of clusters increases, so does WCSS, showing an increase in cluster separation [31]. However, when the graph reaches the elbow point at cluster number 3, the growth ceases to be substantial, suggesting that adding more clusters provides no meaningful benefits in terms of increasing cluster separation. In addition, the best cluster findings were found. The K-Medoids clustering evaluation in Indonesia provides a more accurate assessment of educational quality based on measures of primary, secondary, and senior high school school enrollment rate, gross enrollment rate, and pure participation rates. Table 1 shows the cluster evaluation results.

Table 1 Cluster Evaluation of Education Quality Results in Indonesian Regions

The Dunn's Index	Coefficient of Connectivity	Silhouette Score
0.1651	7.1405	0.2388

The findings of the metric evaluation for K-Medoids clustering based on education quality indicators in various regions of Indonesia are presented in this paper. These metrics include Dunn's Index, Connectivity Coefficient, and Silhouette Score, which provide insight into the clusters' degree of separation, homogeneity, and level of confidence[32].

a. **The Dunn's Index:** Cluster separation is moderate, as indicated by a Dunn's Index of 0.1651. This shows that there is some difference between regions in terms of educational quality, however, it is not significant. The K-Medoids clustering results suggest a moderate level of distinction between regions depending on the quality of education in Indonesia, according to Dunn's score of 0.1651. Even though there is considerable separation, this score shows that the differences between clusters have not yet reached a sufficient level that might be deemed strong separation [33]. This modest separation may indicate that educational quality varies across various places, yet there are still some commonalities or overlaps between groups. As a result, clustering data must be examined in greater depth to identify reasons that underlie variances or similarities in educational quality among regions.
b. **Coefficient of Connectivity**: with a Connectivity Coefficient of 7.1405, cluster homogeneity is moderate. This demonstrates that educational quality is equivalent across geographic groups. The connection coefficient of 7.1405 indicates that cluster homogeneity is moderate. This score suggests that regional groups have a comparable level of educational quality similarity [34]. Even though it is uniform, changes between regions are sufficient to enable clear differentiation between clusters. The occurrence of a certain level of uniformity in the quality of education in each regional group might be regarded as moderate cluster homogeneity. Even though there is some consistency, distinctions between regions provide considerable differences to distinguish the characteristics of each cluster. This enables a clear assessment of the elements influencing educational quality in diverse places.
c. **Silhouette Score**: The silhouette score of 0.2388 indicates a medium level of confidence in clustering observations. Despite some separation across clusters, these findings suggest that grouping confidence has not yet reached a high level. Certain clusters overlap, indicating parallelism or uncertainty in observations. In the context of education quality clusters in various regions of Indonesia, this silhouette score can be viewed as the presence of variances that are not separate within groups. Despite the distinction, several locations share characteristics with other groups [34]. The profile of education quality in Indonesia reveals great variability after clustering using the k-medoids method with an ideal number of clusters of three. Figure 3 depicts the distribution of these cluster results, which provides a visual picture of how provinces lie within each cluster. This research can comprehend the level of diversity in the features of education quality in various regions of Indonesia by looking at the distribution of provinces in each cluster.

Table 2. Characteristics of Each Cluster Based on Education Indicators (Mean Score)

Cluster	APS-SD	APS-SMP	APS-SMA	APK-SD	APK-SMP	APK-SMA	APM-SD	APM-SMP	APM-SMA
Cluster 1	99.35	96.28	72.73	104.65	93.86	83.94	97.72	82.21	62.27
Cluster 2	99.39	97.82	81.71	106.55	93.20	94.44	97.85	81.93	69.63
Cluster 3	97.92	92.85	70.98	106.50	86.54	85.95	96.56	74.96	59.16

Table 2 presents the characteristics of each cluster based on education indicators. This data presents an in-depth view of disparities in educational quality in each regional group as a result of clustering analysis.

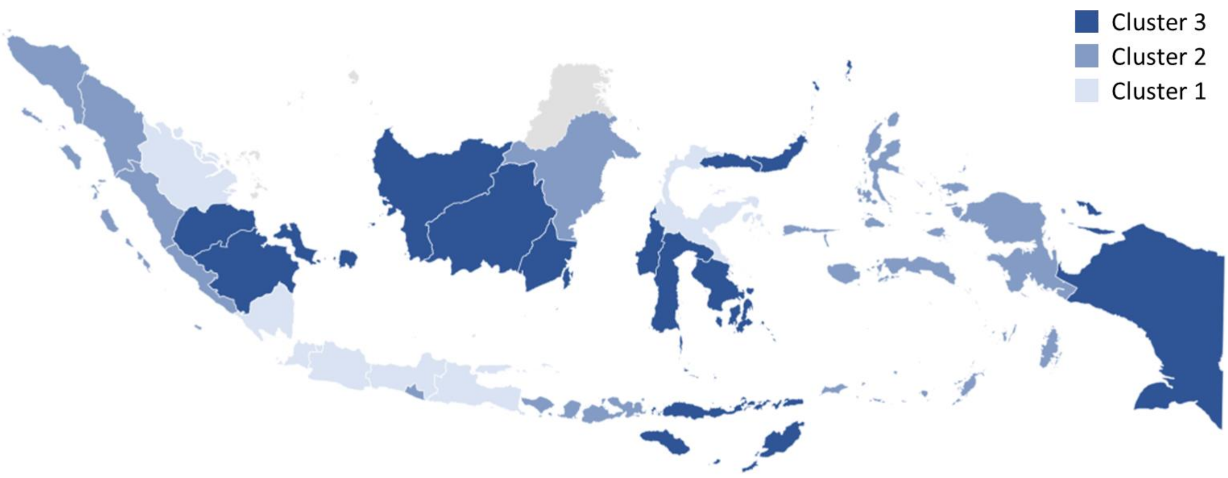

Figure 3. Distribution of Provinces Based on Education Quality Cluster Results in Indonesia
(Source: authors, 2023)

1. **Characteristics of Education Quality in Cluster 1**

Cluster 1. which encompasses eight provinces which includes eight provinces presented in Figure 4 including Riau. Lampung. DKI Jakarta. West Java. Central Java. East Java. Banten. and Central Sulawesi. performs well. particularly in the "APK" and "APM" indicators at the junior high school level. These findings show that these areas have high junior high school enrollment rates. as well as high graduation rates. Furthermore, the pure participation rate in this region is very high, indicating an efficient education system capable of maintaining student engagement levels till junior high school. This advantage paints a favorable picture of efforts to improve the quality of education in this cluster. which can serve as a model for other regions seeking to improve the quality of junior high school education.

Figure 4. Distribution of Provinces in Cluster 1 for Analysis of Education Quality in Indonesia
(Source: authors. 2023)

a. The "APK" indicator in junior high school

This data indicates that 93.86% (in Table 2) of the junior high school-age population in this cluster attends junior high school (SMP). This high percentage demonstrates the community's active participation in enrolling their children in this level of education.

b. The "APM" indicator in junior high school

This data indicates that around 82.21% (in Table 2) of the junior high school age population in this cluster has entered and is enrolled in junior high school. This demonstrates that the majority of the people in this region of junior high school age actively participate in junior high school.

The combination of high "APK" and "APM" indicators at the junior high school level implies that not only do many students participate in junior high school. but the majority of them complete it. This cluster might be viewed as an area that actively supports junior high school education. with significant community participation in enrolling children at this level and a high graduation rate. This can serve as an example for other regions seeking to improve education quality in junior high school.

2. Characteristics of Education Quality in Cluster 2

In addition to demonstrating superior achievement in APS at the Senior High School (SMA) level. this cluster. which comprises 13 provinces presented in Figure 5 including Aceh. North Sumatra. West Sumatra. Bengkulu. Riau Islands. DI Yogyakarta. Bali. West Nusa Tenggara. East Kalimantan. North Kalimantan. Maluku. North Maluku. and West Papua. also dominates other indicators. In this cluster. high school student involvement has attained a high degree. The proportion of high school-age students actively engaged in higher education demonstrates the awareness and dedication of society towards this goal. The region's high APS-SMA may be seen as a sign of public understanding of the value of higher education in raising the caliber of human resources. It has been successful for the local government and education partners to establish an atmosphere that encourages high school student involvement.

Figure 5. Distribution of Provinces in Cluster 2 for Analysis of Education Quality in Indonesia
(Source: authors. 2023)

This area. with an APS-SMA of 81.71% (in Table 2). can serve as a model for initiatives aimed at raising high school student participation rates by highlighting the significance of public awareness and government support for postsecondary education. More chances for innovation. human resource development. and raising people's standards of living are presented by higher education. The provinces in Cluster 2 serve as excellent examples of how to encourage student involvement in postsecondary education. The high APS-SMA shows that both the general population and the local government recognize the value of higher education and are working to develop infrastructure and facilities that will support it. This achievement may serve as an example for other areas looking to boost high school student engagement and broaden the public's access to higher education.

3. Characteristics of Education Quality in Cluster 3

The provinces that make up this cluster presented in Figure 6. which exhibits the highest degree of performance. are Jambi. South Sumatra. Bangka Belitung Islands. East Nusa Tenggara. West Kalimantan. Central Kalimantan. South Kalimantan. North Sulawesi. South Sulawesi. Southeast Sulawesi.

Gorontalo. West Sulawesi. and Papua. low in terms of educational metrics. ranging from high school (SMA) to elementary school (SD). Low graduation rates from elementary to high school present significant issues for this cluster. as does expanding student involvement at the primary and secondary education levels. Low achievement is also reflected in the APK for grades elementary through senior high school. This low graduation rate suggests that challenges exist in meeting the objective of a 12-year education mandate in this area.

Figure 6. Distribution of Provinces in Cluster 3 for Analysis of Education Quality in Indonesia
(Source: authors. 2023)

According to the pure participation rate (APM) from elementary to senior high school, in comparison to other clusters, this cluster shows the lowest percentage of school-aged children attending certain levels of education. This statistic gives an idea of the proportion of school-age children in a region that attends a particular level of education. Cluster 3's troubles continue beyond high school and influence all educational levels. As a result. a comprehensive and well-coordinated approach is required to increase the standard of education in this sector from elementary to high school. Careful planning and teamwork are essential to overcome these challenges and ensure that every community in Cluster 3 has access to better. more sustainable schooling.

4. CONCLUSION

This study successfully identified three main clusters of educational quality in Indonesia using the k-medoids algorithm, based on indicators such as school enrollment rate, gross enrollment rate, and pure participation rate at the elementary, junior high, and senior high school levels. The machine learning approach provided a clearer map of educational quality, enabling more specific and effective education policies. However, the study faced limitations due to the data's scope and quality, and the k-medoids algorithm's efficiency with very large datasets. Additionally, the study only considered a few educational indicators. Future developments could involve using more extensive and detailed datasets, combining k-medoids with other algorithms, exploring additional external factors, and implementing the study's findings in real policy-making.

ACKNOWLEDGMENT

The researchers extend their gratitude for the support provided by the Study Program in Informatics Engineering, Faculty of Computer Science at Universitas Dian Nuswantoro in this study.

REFERENCES

[1] Suharno, N. A. Pambudi, and B. Harjanto, "Vocational education in Indonesia: History, development, opportunities, and challenges," *Child Youth Serv Rev*, vol. 115, p. 105092, Aug. 2020, doi: 10.1016/J.CHILDYOUTH.2020.105092.

[2] S. J. Daniel, "Education and the COVID-19 pandemic," *Prospects (Paris)*, vol. 49, no. 1–2, pp. 91–96, Oct. 2020, doi: 10.1007/S11125-020-09464-3/METRICS.

[3] M. Nurtanto, N. Kholifah, A. Masek, P. Sudira, and A. Samsudin, "Crucial Problems in Arranged the Lesson Plan of Vocational Teacher.," *International Journal of Evaluation and Research in Education*, vol. 10, no. 1, pp.

345–354, Mar. 2021, doi: 10.11591/ijere.v10i1.20604.
[4] N. Permatasari and A. Ubaidillah, "Estimation of Education Indicators in East Java Using Multivariate Fay-Herriot Model," *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2021, no. 1, pp. 108–118, Jan. 2021, doi: 10.34123/ICDSOS.V2021I1.51.
[5] S. Hanifah and A. H. Primandari, "Implementasi Metode K-Means Clustering dalam Pengelompokan Kabupaten/ Kota di Provinsi NTB Berdasarkan Indikator Pendidikan," *Emerging Statistics and Data Science Journal*, vol. 1, no. 3, pp. 378–393, Dec. 2023, doi: 10.20885.10.
[6] Z. HASAN, "The Effect of Human Development Index and Net Participation Rate on the Percentage of Poor Population: A Case Study in Riau Province, Indonesia," *International Journal of Islamic Economics and Finance Studies*, vol. 8, no. 1, pp. 24–40, Mar. 2022, doi: 10.54427/IJISEF.964861.
[7] K. A. Azhar, N. Iqbal, Z. Shah, and H. Ahmed, "Understanding high dropout rates in MOOCs – a qualitative case study from Pakistan," *Innovations in Education and Teaching International*, Apr. 2023, doi: 10.1080/14703297.2023.2200753.
[8] C. Lowder, C. O'Brien, D. Hancock, J. Hachen, and C. Wang, "High School Success: A Learning Strategies Intervention to Reduce Drop-Out Rates," *Urban Review*, vol. 54, no. 4, pp. 509–530, Nov. 2022, doi: 10.1007/S11256-021-00624-Z/TABLES/4.
[9] B. Lund and J. Ma, "A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering," *Performance Measurement and Metrics*, vol. 22, no. 3, pp. 161–173, Nov. 2021, doi: 10.1108/PMM-05-2021-0026/FULL/XML.
[10] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/S42979-021-00592-X/FIGURES/11.
[11] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," *IEEE Access*, vol. 8, pp. 151847–151855, 2020, doi: 10.1109/ACCESS.2020.3014021.
[12] T. Widiyaningtyas, M. I. W. Prabowo, and M. A. M. Pratama, "Implementation of k-means clustering method to distribution of high school teachers," *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, vol. 2017-December, Dec. 2017, doi: 10.1109/EECSI.2017.8239083.
[13] S. K. Dini and A. Fauzan, "Clustering Provinces in Indonesia based on Community Welfare Indicators," *EKSAKTA: Journal of Sciences and Data Analysis*, vol. 1, no. 1, pp. 56–63, Feb. 2020, doi: 10.20885/EKSAKTA.VOL1.ISS1.ART9.
[14] M. R. Putri, G. S. Nugraha, and R. Dwiyansaputra, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering:," *Journal of Computer Science and Informatics Engineering (J-Cosine)*, vol. 7, no. 1, pp. 76–83, Jun. 2023, doi: 10.29303/JCOSINE.V7I1.509.
[15] M. Monica, N. U. Ayuningtiyas, H. Al Azies, M. Riefky, H. Khusna, and S. P. Rahayu, "Unsupervised Learning Approach for Evaluating the Impact of COVID-19 on Economic Growth in Indonesia," *Communications in Computer and Information Science*, vol. 1489 CCIS, pp. 54–70, 2021, doi: 10.1007/978-981-16-7334-4_5/COVER.
[16] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 8, pp. 10173–10196, Aug. 2023, doi: 10.1109/TPAMI.2023.3250241.
[17] A. V. Ushakov and I. Vasilyev, "Near-optimal large-scale k-medoids clustering," *Inf Sci (N Y)*, vol. 545, pp. 344–362, Feb. 2021, doi: 10.1016/J.INS.2020.08.121.
[18] F. Zahra, A. Khalif, B. N. Sari, U. S. Karawang, J. H. Ronggo Waluyo, and T. Timur, "PENGELOMPOKAN TINGKAT KEMISKINAN DI SETIAP PROVINSI DI INDONESIA MENGGUNAKAN ALGORITMA K-MEDOIDS," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, pp. 2830–7062, Apr. 2024, doi: 10.23960/jitet.v12i2.4199.
[19] A. J. Onumanyi, D. N. Molokomme, S. J. Isaac, and A. M. Abu-Mahfouz, "AutoElbow: An Automatic Elbow Detection Method for Estimating the Number of Clusters in a Dataset," *Applied Sciences 2022, Vol. 12, Page 7515*, vol. 12, no. 15, p. 7515, Jul. 2022, doi: 10.3390/APP12157515.
[20] T. M. Murugan and E. Baburaj, "Clustering and classification with inertia weight and elitism-based particle swarm optimization," *Pattern Analysis and Applications*, vol. 24, no. 4, pp. 1605–1621, Jul. 2021, doi: 10.1007/S10044-021-01010-X/FIGURES/5.
[21] J. Y. Kim, G. Park, S. A. Lee, and Y. Nam, "Analysis of Machine Learning-Based Assessment for Elbow Spasticity Using Inertial

Sensors," *Sensors 2020, Vol. 20, Page 1622*, vol. 20, no. 6, p. 1622, Mar. 2020, doi: 10.3390/S20061622.

[22] B. W. Otok, A. Suharsono, Purhadi, R. E. Standsyah, and H. Al Azies, "Partitional Clustering of Underdeveloped Area Infrastructure with Unsupervised Learning Approach: A Case Study in the Island of Java, Indonesia," *Journal of Regional and City Planning*, vol. 33, no. 2, pp. 177–196, Aug. 2022, doi: 10.5614/JPWK.2022.33.2.3.

[23] T. Gupta and S. P. Panda, "Clustering Validation of CLARA and K-Means Using Silhouette DUNN Measures on Iris Dataset," *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, pp. 10–13, Feb. 2019, doi: 10.1109/COMITCON.2019.8862199.

[24] H. Habiballoh, A. Faqih, and T. Suprapti, "IMPLEMENTASI ALGORITMA K-MEANS DALAM MENGELOMPOKAN KABUPATEN/KOTA DI JAWA BARAT BERDASARKAN JENIS DAN JUMLAH POTENSI OBJEK DAYA TARIK WISATA," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, pp. 2830–7062, Apr. 2024, doi: 10.23960/JITET.V12I2.4270.

[25] G. Ranganathan, "Hyperspectral Image Processing in Internet of Things model using Clustering Algorithm," *Journal of ISMAC*, vol. 03, no. 02, pp. 163–175, 2021, doi: 10.36548/jismac.2021.2.008.

[26] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," *Entropy 2021, Vol. 23, Page 759*, vol. 23, no. 6, p. 759, Jun. 2021, doi: 10.3390/E23060759.

[27] F. Salsabila, T. Ridwan, U. Singaperbangsa Karawang, J. HSRonggo Waluyo, and T. Timur, "ANALISA VOLUME PENYEBARAN SAMPAH DI KARAWANG MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, pp. 2830–7062, Apr. 2024, doi: 10.23960/JITET.V12I2.4226.

[28] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, pp. 747–748, Oct. 2020, doi: 10.1109/DSAA49011.2020.00096.

[29] SchubertErich, "Stop using the elbow criterion for k-means and how to choose the number of clusters instead," *ACM SIGKDD Explorations Newsletter*, vol. 25, no. 1, pp. 36–42, Jun. 2023, doi: 10.1145/3606274.3606278.

[30] O. Pasin and S. Gonenc, "An investigation into epidemiological situations of COVID-19 with fuzzy K-means and K-prototype clustering methods," *Scientific Reports 2023 13:1*, vol. 13, no. 1, pp. 1–11, Apr. 2023, doi: 10.1038/s41598-023-33214-y.

[31] K. Gratsos, S. Ougiaroglou, and D. Margaris, "A Web Tool for K-means Clustering," *Lecture Notes in Networks and Systems*, vol. 783 LNNS, pp. 91–101, 2023, doi: 10.1007/978-3-031-44097-7_9/COVER.

[32] M. Z. Rodriguez *et al.*, "Clustering algorithms: A comparative approach," *PLoS One*, vol. 14, no. 1, p. e0210236, Jan. 2019, doi: 10.1371/JOURNAL.PONE.0210236.

[33] T. Olivoto and A. D. C. Lúcio, "metan: An R package for multi-environment trial analysis," *Methods Ecol Evol*, vol. 11, no. 6, pp. 783–789, Jun. 2020, doi: 10.1111/2041-210X.13384.

[34] G. Brock, V. Pihur, S. Datta, and S. Datta, "clValid: An R Package for Cluster Validation," *J Stat Softw*, vol. 25, no. 4, pp. 1–22, Mar. 2008, doi: 10.18637/JSS.V025.I04.