ICMEM

The 9th International Conference on Management in Emerging Markets

Region-Based Recommendations for Upper Secondary School Types in Indonesia Using Clustering Analysis

Evan Haryowidyatna[1]*, Manabu Ichikawa[1], Kaede Fujita[1]

[1]Shibaura Institute of Technology, Saitama, Japan

Abstract *- Indonesia faces significant inequality in access to education between urban and rural areas, made more complex by its geographic and cultural diversity. With the government allowing students to choose between general and vocational upper secondary schools, it is crucial to guide them toward the most suitable path to support their future careers and ensure they can thrive in their own regions. This study analyzes all 514 Indonesian cities and regencies using socioeconomic data from 2019 to 2024 to identify appropriate education pathways. Clustering methods K-Means and K-Medoids were applied, with K-Means using five clusters producing the best results. Cluster 0 (18 regions) and Cluster 2 (5 regions) are better suited for general education due to high technology use and rapid industry growth. Cluster 1 (211 regions) and Cluster 3 (194 regions) are more aligned with vocational education, reflecting lower technology use and slower development. Cluster 4 (86 regions) has mixed characteristics. These findings can help shape education policy to match local conditions. Future research should include cultural and community-specific factors to further improve education equity across Indonesia.*

Keywords *- Clustering Analysis, Education, Evaluation Method, Socioeconomic Indicators, Upper Secondary School*

I. INTRODUCTION

Education plays a crucial role in personal development and the progress of society. However, significant differences in access and quality still exist between urban and rural areas, limiting learning opportunities for people in remote regions. In rural areas, poor access to quality education often results in a shortage of skilled workers, which slows down local economic growth. In contrast, urban areas generally have better educational infrastructure and outcomes [1]. These challenges are further intensified by Indonesia's vast geography, consisting of more than 17,000 islands [2] and over 700 regional languages [3], making it a complex task to provide equal and inclusive education across the country.

To meet the educational needs across the country, the Indonesian government enacted Law Number 20 of 2003 on the National Education System. This law defines formal education in Indonesia as six years of elementary school, followed by three years of lower secondary school, and another three years of upper secondary school. At the upper secondary level, students can choose between two pathways: general (academic) or vocational. The general pathway prepares students for university, whereas the vocational pathway focuses on practical job skills for those planning to enter the workforce directly [4]. Given this option, it is important to identify which type of school is most suitable for each region. Since both general and vocational schools play valuable roles in society, more effort is needed to ensure that students are guided toward the most appropriate path starting from upper secondary school.

There are two key theoretical perspectives that support the need to guide students toward the most suitable type of schooling: Sen's Capability Approach [5] and Bourdieu's Theory of Social and Cultural Capital [6]. The capability approach emphasizes that educational inequality is not only about access but also about ensuring individuals have real freedom to pursue educational paths aligned with their goals and local conditions. Bourdieu's theory highlights how family background, cultural norms, and access to resources shape students' educational choices. In rural areas, limited exposure to academic culture and fewer resources often lead students to view school mainly as a route to employment, making vocational education

*Corresponding author. Email: mf24502@shibaura-it.ac.jp

more appealing. In contrast, students in urban areas typically benefit from stronger academic support networks and may view education as a broader tool for personal and social development. These frameworks underscore the importance of tailoring education policies to reflect regional differences rather than applying a uniform model across the country.

Although vocational schools help students, especially in rural areas [7] and to enter the workforce more quickly [8], relying too much on vocational education may create a workforce with limited flexibility in a rapidly changing job market [9]. This raises an important question: do the types of education offered in each region match the local conditions? Understanding the diverse characteristics of each region, such as economic activity, infrastructure, and population trends, is essential for aligning education strategies with regional needs.

To address this issue, this research introduces a new approach by clustering all cities and regencies in Indonesia based on socioeconomic indicators. The aim is to determine the most suitable type of education, either general or vocational, for each area, focusing on improving outcomes at the upper secondary level. This study identifies key regional groupings, determines the optimal number of clusters, and visualizes the results through map to support equitable, region-specific education policies. By analyzing data from all cities and regencies in Indonesia, the research captures the nation's rich diversity and offers a robust foundation for developing more targeted and inclusive education strategies at the national level.

II. LITERATURE REVIEW

Indonesia is a culturally and geographically diverse country [3]. Educational outcomes are shaped by various factors, including social and cultural backgrounds, financial conditions, teaching methods, school types, infrastructure, and the transition from school to employment. In remote regions such as Papua [10] and Sulawesi [11], place based education that involves indigenous teachers and students has been shown to improve learning outcomes by fostering respect for local culture. Financial conditions also play a significant role; for example, studies in Aceh and North Sulawesi found that parental financial support enhances academic performance [12], with other research indicates that economic hardship can force children to work to support their families [13]. These findings underscore the importance of considering the socioeconomic context when addressing educational challenges.

Another key factor is the transition from education to employment. A study conducted in Jakarta, Indonesia's capital, highlights the importance of strengthening the education system to better support students in securing stable jobs [14]. Vocational high schools provide clear benefits by helping students enter the workforce earlier than graduates from general academic schools [8]. However, while vocational education offers practical advantages, relying too heavily on it may result in a workforce with limited adaptability due to narrowly focused skills [9]. This highlights the need for deeper analysis to determine the most suitable type of upper secondary school for each region to improve educational outcomes.

To address the diversity of regions in Indonesia and recommend the most suitable type of school, clustering analysis has recently gained popularity as a method for studying cities and regencies. In the education sector, clustering has shown that the quality of elementary school inputs and outcomes in Indonesia is not necessarily linked to a region's distance from the capital [15]. It has also identified at least 14 distinct school types in West Java based on the availability of educational facilities [16]. Beyond academic research, clustering also plays an important role in shaping government priorities and guiding economic policies [17].

In addition to clustering analysis, socioeconomic factors are essential for identifying meaningful regional groupings in Indonesia. In the education sector, recent studies have confirmed that socioeconomic status strongly influences student outcomes, and that education policies must consider various interconnected factors rather than applying a one-size-fits-all solution [18]. These findings highlight that combining clustering analysis with socioeconomic data can provide valuable insights for developing more targeted and effective education strategies.

International research also supports the value of regional clustering in education planning. In China, clustering has been used to analyze the spatial distribution of education quality, helping with better resource allocation [19]. Specifically, clustering using Fuzzy C-Means identified disparities in teacher distribution across 31 provinces, leading to targeted improvements in compulsory education [20]. These global examples show that data-driven regional planning can promote more equitable and localized education strategies, reinforcing the relevance of this method for Indonesia.

Although previous studies have offered valuable insights into improving education in Indonesia, most focus on specific regions, such as Papua or Jakarta, or emphasize identifying ideal school types without considering which areas need them the most. This often leads to recommendations that may not be suitable for all regions. To our knowledge, there remains a gap in research that examines which types of schools are most appropriate for every city and regency in Indonesia, based on their unique characteristics. This study seeks to address that gap by applying clustering analysis to socioeconomic data in order to recommend more suitable and targeted upper secondary educational approaches for each region.

III. METHODOLOGY

A. Data

The data used in this study were sourced from Statistics Indonesia (BPS), which is the most appropriate source based on Law No. 16 of 1997 on Statistics [21]. According to this law, BPS is responsible for providing data for the government and the public through censuses, surveys, and secondary data from other government institutions. This study uses data from all 514 cities and regencies in Indonesia, covering the years 2019 to 2024. The average value across these years is used in the analysis. Eleven variables representing socioeconomic indicators were selected from BPS, including Gross Regional Domestic Product at Constant Prices, Poverty Line, Average Years of Schooling, Average Monthly Expenses per Capita, Minimum Monthly Wage, Percentage of Inter-Village Roads Passable by Four-Wheeled Vehicles, Population Density, Percentage of Households Owning Motorcycles, and the number of people using computers, mobile phones, and the internet.

B. Clustering

In this study, the partitional clustering method was selected. Partitional clustering is considered more suitable for handling large datasets compared to hierarchical methods, and it enables more efficient analysis when working with larger samples. However, this method also has some limitations. It is sensitive to outliers, the selection of irrelevant or inappropriate variables, and the choice of distance measures used in the analysis [22]. Two commonly used partitional clustering methods applied in this study are K-Means and K-Medoids [23]. To determine which method produces the best clustering results, three internal evaluation metrics are used: the silhouette index, the Davies Bouldin index, and the Dunn index. The best clustering method is chosen based on the consistency of these evaluation results.

C. K-Means

The K-Means Clustering algorithm uses the centroid of a cluster as the average of all objects within that cluster. It applies a partitioning method to divide data into several groups, where each member in a group shares similar characteristics with others in the same group [24]. Below is the algorithm for performing K-Means clustering.

1. Set the number of clusters $\boldsymbol{k}$ to determine how many groups the data will be divided into.
2. Determine the initial value of the centroid or the center point of the cluster randomly. Recalculate the position of each cluster's centroid by taking the mean of all data points currently assigned to it. This new centroid becomes the center point for the cluster in the next iteration using the following equation:

$$\overline{\mathrm{V}}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \tag{1}$$

   Let $\overline{\mathrm{V}}_{ij}$ represent the centroid of the i-th cluster for the j-th variable, where N_i is the number of data points in the i-th cluster. The index k refers to the cluster number, j refers to the variable index, and X_{kj} is the value of the k-th data point in the cluster for the j-th variable.
3. Calculate the distance between each data point and the centroids using the Euclidean distance formula.

$$d_{e(x,y)} = \sqrt{\sum_{i=1}^{n} \left(x_{ik} - x_{jk}\right)^2} \tag{2}$$

   Let $d_{e(x,y)}$ denote the distance between the i-th and j-th objects. The variable n represents the number of data points in the i-th cluster. The term x_{ik} refers to the i-th data object among k data points, while x_{jk} refers to the j-th data object in the k-th dataset, which is often considered the centroid (or midpoint value) of the cluster for the total number of data points n.
4. Assign each data point to the cluster with the nearest centroid, based on the minimum distance.
5. Update the centroid of each cluster by recalculating the average position of all data points within that cluster.
6. Repeat Step 2 to the end until the centroid value no longer changes.

D. K-Medoids

The difference between K-Means and K-Medoids is that K-Means uses the mean of points as the cluster center, whereas K-Medoids uses actual data points (medoids) as centers. K-Medoids is more robust to outliers compared to K-Means, which is sensitive to outliers because they can be located far from the cluster's average [23]. Below is the algorithm for performing K-Medoids clustering.

1. Initialize k cluster centers (k = number of clusters).
2. Calculate the nearest cluster for each object using Euclidean distance as in Equation (2).
3. After calculating the distances, randomly select new cluster centers from the data as non-medoid candidates.

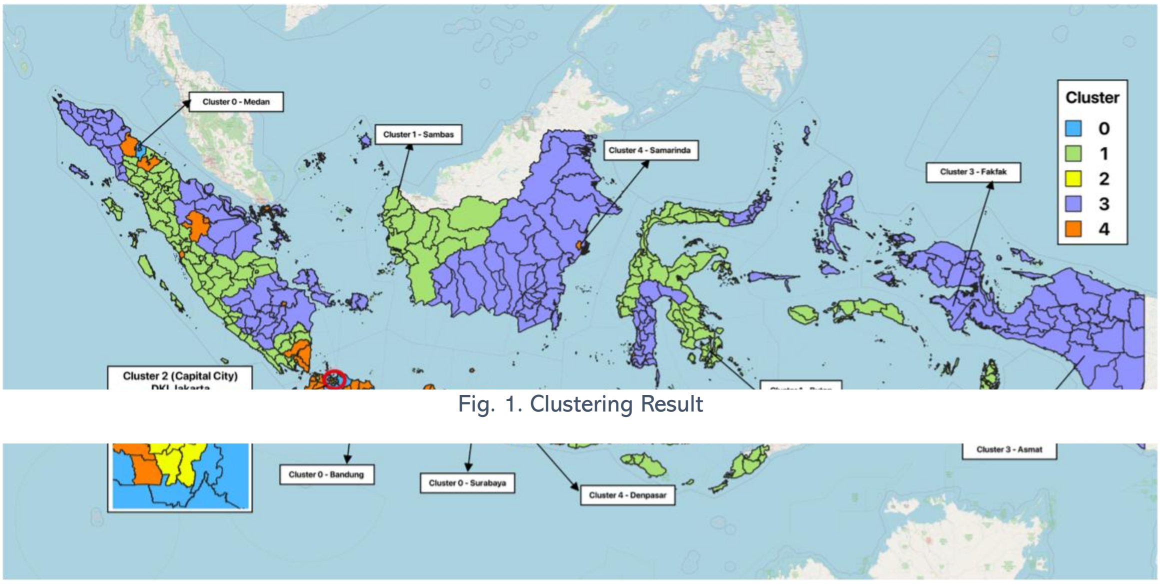

Fig. 1. Clustering Result

4. Compute the distance from each object in a group to the non-medoid candidates.
5. Calculate the total deviation (S) by subtracting the old total distance from the new total distance. If $S < 0$, swap the current medoid with the non-medoid candidate to form a new set of K-Medoids.
6. Repeat Steps 3 to 5 until the medoids no longer change, resulting in the final clusters and their members.

E. Evaluation Method

Cluster evaluation can be performed using external methods with labels or internal methods that assess quality based on the data itself. In this study, internal evaluation is more appropriate. The internal evaluation methods used are the Silhouette Index, Davies-Bouldin Index, and Dunn Index. The target of using multiple evaluation method is checking the consistency of result.

1. The Silhouette Index (SI) is one of the most popular and effective methods for evaluating clustering results [25]. It measures the average distance between a data point and others in the same cluster compared to those in other clusters. Scores range from -1 to 1, with higher positive values indicating better clustering. Negative values suggest poor clustering, where points are closer to other clusters than their own [26].
2. The main goal of the Davies-Bouldin Index (DBI) is to maximize the distance between clusters (inter-cluster distance) while minimizing the distance between data points within the same cluster (intra-cluster distance). The smaller the DBI value, the better the clustering result [27].
3. Dunn's index (DI) validates clustering by comparing the distance between clusters to the cluster diameters. It is the ratio of the minimum inter-cluster distance to the maximum intra-cluster diameter. Higher values indicate better, well-separated, and compact clusters [28].

IV. FINDINGS AND DISCUSSION

A. Clustering Result

Based on the clustering evaluation metrics applied to both K-Means and K-Medoids, Table 1 indicates that K-Means clustering with five clusters provides the most optimal results. This conclusion is supported by two out of three evaluation measures where Silhouette Coefficient and Davies-Bouldin Index identify K-Means with five clusters as the best-performing method for classifying Indonesian cities and regencies. The consistency across these metrics strengthens the credibility of this result.

The five-cluster result shows a varied distribution of cities and regencies, with cluster sizes of 18, 211, 5, 194, and 86. As shown in Fig. 1, most cities in the Island of Java fall into the orange cluster, highlighting a distinct pattern compared to other islands, which are mainly in green and purple. This suggests that the characteristics of cities and regencies in Java differ notably from those in other regions. Notably, DKI Jakarta forms its own cluster, underscoring its unique socioeconomic and demographic profile.

Table I Cluster Evaluation Result

	K-Means			K-Medoids		
	4	5	6	4	5	6
SI	0.39	0.40 (Best)	0.38	-0.21	-0.29	-0.22
DI	0.01	0.01	0.01	0.01	0.03	0.06 (Best)
DBI	0.88	0.80 (Best)	0.99	0.88	0.91	1.01

B. Socioeconomic Characteristics of Clusters in Indonesia

The five clusters show clear socioeconomic and infrastructural distinctions. Cluster 0 includes major, densely populated cities like Surabaya, Bandung, and Medan, with high internet and mobile phone usage, good road infrastructure, strong GDP and education, average poverty, but the lowest minimum wage. Cluster 1 covers smaller, less-known regions such as Sambas and Buton, with low infrastructure, minimal tech access, and the lowest expenses and GDP, though with average wages and education. Cluster 2 consists solely of DKI Jakarta, showing the strongest economy, highest computer use and road quality, high population and connectivity, and moderate vehicle ownership. Cluster 3 includes the least developed and least populous areas like Fakfak and Asmat, mainly outside the Island of Java, with low infrastructure, low GDP, and low education, yet high wages and poverty. Cluster 4, which includes cities like Samarinda, Denpasar, and most on the Island of Java, possesses a distinct socioeconomic profile with very high motorcycle ownership, low expenses and wages, and the lowest poverty rate, alongside moderate infrastructure and an average population.

C. Discussion

Choosing between vocational and academic education remains a significant challenge in Indonesia. Although vocational schools offer clear benefits by helping students enter the workforce more quickly [8], this approach may also produce a workforce that is overly specialized and less adaptable to evolving industry demands [9]. Vocational education is often more suitable for rural regions [7], which tend to have lower income levels, limited access to services, and fewer educational resources [29]. However, relying too heavily on vocational pathways could limit long-term flexibility and economic mobility, particularly in a dynamic labor market.

Based on the clustering results, Cluster 0 and Cluster 2 are regions characterized by higher levels of technological adoption and infrastructure which better positioned to benefit from an academic education focus. In these areas, educational strategies could prioritize expanding STEM-based programs, university preparation courses, and the integration of digital learning tools to help students develop critical thinking and transferable skills. Strengthening academic pathways in these clusters could support local innovation and align with fast-paced industrial development.

Conversely, Cluster 1 and Cluster 3 show slower industrial growth and limited access to technology, suggesting a stronger alignment with vocational education. In these regions, vocational curricula should be tailored to local economic conditions. For instance, in areas dependent on agriculture or fisheries, vocational training could focus on modern farming methods, agribusiness, or aquaculture. Local governments and schools should collaborate with small and medium-sized enterprises to provide hands-on apprenticeships and industry-linked internships that directly respond to local labor market needs. Such partnerships can ensure that students are equipped with relevant, employable skills, thereby supporting regional economic development.

Cluster 4 appears to be in a transitional stage between high and low-development regions. Here, a dual-track model combining academic and vocational elements may be the most effective strategy. Schools in these areas could offer flexible programs that allow students to pursue both academic subjects and job-related skills, empowering them to either continue to higher education or enter the workforce depending on their preferences and opportunities.

To ensure successful implementation, policymakers and school administrators should take several key actions. First, they should conduct regional labor market assessments to better align educational offerings with actual employment demands. Second, investment in teacher training is crucial to improve the delivery of specialized vocational content. Third, partnerships with local industries must be strengthened to keep curricula up to date and provide real-world experience for students. Governments could also introduce incentives such as tax relief or public recognition to encourage businesses to engage in vocational training programs.

Lastly, future research should explore more granular data at the district or village level, incorporating physical geography, such as mountainous or coastal terrains, to refine educational planning. This study confirms prior findings that vocational education can benefit less developed areas [7] but adds more specific, data-driven insights on which regions should prioritize which education types.

V. CONCLUSION

This study addresses the challenge of aligning education pathways with Indonesia's diverse regional conditions, characterized by significant socioeconomic disparities. Using socioeconomic data from 514 cities and regencies, partitional clustering methods were applied, with K-Means producing the most effective classification into five distinct clusters: Cluster 0 (18 regions), Cluster 1 (211 regions), Cluster 2 (5 regions), Cluster 3 (194 regions), and Cluster 4 (86 regions). The clusters reveal clear differences in economic development, infrastructure, and digital connectivity. Based on these profiles, academic education is recommended for the more developed clusters, Cluster 0 and Cluster 2, which have high levels of technology use that may lead to rapidly changing industries. In contrast, vocational education is better suited for the less developed clusters, Cluster 1 and Cluster 3, due to their low technology use, suggesting that industries there may not change quickly. Cluster 4, with transitional characteristics, may benefit from a combined approach. These insights provide a data-driven foundation for targeted education policies.

Although this research provides a strong foundation for developing targeted education policies, it has several limitations. One key limitation is the diverse geographical landscape of Indonesia, which ranges from coastal areas to mountainous regions and may influence access to education and infrastructure needs. The study also does not account for local cultural factors that can shape community attitudes and preferences toward education. Future research should examine these cultural and social dynamics, including how vocational schools contribute to building interpersonal relationships and community networks. Additionally, geographical elements such as elevation and transportation accessibility play a crucial role in determining the suitability of online learning versus traditional classroom settings. To create more effective and inclusive education strategies, further analysis at the district or village level is essential to capture the specific needs and conditions of each local area.

ACKNOWLEDGMENT

We sincerely thank Professor Manabu Ichikawa of Shibaura Institute of Technology for his invaluable guidance and support throughout this research. We also extend our gratitude to the members of the Ichikawa Laboratory for their insightful advice and kind collaboration.

REFERENCES

[1] Y. Chen, “Differences of Primary Education Resources between Urban and Rural Areas,” International Journal of Education and Humanities, vol. 17, no. 2, pp. 223–225, Nov. 2024, doi: 10.54097/7w9qq881.

[2] M. Widyartono and R. Rahmadian, "OTEC Potential of East Nusa Tenggara Province in Indonesia," IOP Conference Series: Materials Science and Engineering, vol. 336, no. 1, p. 012005, Apr. 2018, doi: 10.1088/1757-899x/336/1/012005.

[3] T. B. Pepinsky, M. Abtahian, and A. Cohn, “Urbanization, Ethnic Diversity, and Language Shift in Indonesia,” SSRN Electronic Journal, Mar. 2020, doi: 10.2139/ssrn.3529422.

[4] Republic of Indonesia, Law No. 20 of 2003 on the National Education System, Jakarta, Indonesia, 2003.

[5] C. R. Beitz, “Development as Freedom, Amartya Sen (New York: Alfred A. Knopf, 1999), 380 pp., $27.50 cloth.,” Ethics & International Affairs, vol. 14, pp. 149–151, Mar. 2000, doi: 10.1017/s0892679400008728.

[6] P. Bourdieu, “The Forms of Capital,” in Handbook of Theory and Research for the Sociology of Education, J. G. Richardson, Ed. New York: Greenwood Press, 1986, pp. 241–258.

[7] M. A. Arham, F. H. Y. Akib, and H. Anam, “Does the education sector contribute to overcoming poverty in the Tomini Bay Area of Indonesia? Method of moments quantile approach,” Int. J. Econ. Finance Stud., vol. 14, no. 4, pp. 114–134, 2022, doi: 10.34111/ijefs.20220107.

[8] K. Ariansyah et al., “Comparing labor market performance of vocational and general school graduates in Indonesia: insights from stable and crisis conditions,” Empirical Research in Vocational Education and Training, vol. 16, no. 1, Apr. 2024, doi: 10.1186/s40461-024-00160-6.

[9] D. Pritadrajati, N. Merdikawati, S. C. Saxena, and A. M. Tjahjadi, “Minimum wage and educational pathways in Indonesia: general or vocational tracks?” Asian Development Review, vol. 41, no. 02, pp. 107–135, Jul. 2024, doi: 10.1142/s0116110524400110.

[10] M. Sianturi, C.-L. Chiang, and A. A. Hurit, “Impact of a Place-Based Education Curriculum on Indigenous teacher and students,” International Journal of Instruction, vol. 11, no. 1, pp. 311–328, Dec. 2017, doi: 10.12973/iji.2018.11122a.

[11] “Education for all: A study on education for indigenous people in South Sulawesi, Indonesia,” Kasetsart Journal of Social Sciences, vol. 42, no. 3, Jan. 2021, doi: 10.34044/j.kjss.2021.42.3.25.

[12] G. Van Der Werf, B. Creemers, and H. Guldemond, “Improving parental involvement in primary education in Indonesia: implementation, effects and costs,” School Effectiveness and School Improvement, vol. 12, no. 4, pp. 447–466, Dec. 2001, doi: 10.1076/sesi.12.4.447.3444.

[13] A. Suryahadi, A. Priyambada, and S. Sumarto, “Poverty, School and Work: Children during the Economic Crisis in Indonesia,” Development and Change, vol. 36, no. 2, pp.

351–373, Mar. 2005, doi: 10.1111/j.0012-155x.2005.00414.x.

[14] M. U. Absor and I. D. Utomo, “School-to-work transition of young people in Greater Jakarta: the determinants of success,” Asian Population Studies, vol. 13, no. 3, pp. 306–326, Aug. 2017, doi: 10.1080/17441730.2017.1351686.

[15] F. Wijayanto, "Indonesia education quality: Does distance to the capital matter? (A clustering approach on elementary school intakes and outputs qualities)," 2015 International Conference on Science and Technology (TICST), Pathum Thani, Thailand, 2015, pp. 318-322, doi: 10.1109/TICST.2015.7369378.

[16] M. A. Putri and S. Abdullah, “Clustering analysis of senior high school in West Java based on educational facilities,” Journal of Physics Conference Series, vol. 1725, no. 1, p. 012032, Jan. 2021, doi: 10.1088/1742-6596/1725/1/012032.

[17] R. Gustriansyah, J. Alie, A. Sanmorino, R. Heriansyah, and M. N. M. M. Noor, “Machine learning for clustering Regencies-Cities based on inflation and poverty rates in Indonesia,” Indonesian Journal of Information Systems, vol. 5, no. 1, pp. 64–73, Aug. 2022, doi: 10.24002/ijis.v5i1.5682.

[18] C. Y. Tan, “Socioeconomic Status and Student Learning: Insights from an Umbrella Review,” Educational Psychology Review, vol. 36, no. 4, Aug. 2024, doi: 10.1007/s10648-024-09929-3.

[19] L. L. Y. C. Yan Li Yang, Ke, “Evaluation of Teacher Allocation in Regional Compulsory Education Industry Based on K-means clustering and GIS Spatial analysis: A Case Study of teacher-student ratio,” CONVERTER, pp. 902–912, Jul. 2021, doi: 10.17762/converter.358.

[20] Z. Li, S. Chen, L. Pei, J. Chu, and J. Song, “Teacher allocation and evaluation based on fuzzy C-Means clustering,” Mobile Information Systems, vol. 2022, pp. 1–9, Jun. 2022, doi: 10.1155/2022/8465713.

[21] R. Yuliarto, M. Zeinudin, and M. Warka, “Optimizing the Organization of Statistical Activities through the Revision of Law No. 16 of 1997 Concerning Statistics,” Journal of Progressive Law and Legal Studies, vol. 1, no. 03, pp. 197–205, Nov. 2023, doi: 10.59653/jplls.v1i03.333.

[22] B. W. Otok, A. Suharsono, P. 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”, jrcp, vol. 33, no. 2, pp. 177-196, Aug. 2022.

[23] J. Han, M. Kamber, and J. Pei, “Cluster analysis: Basic concepts and methods,” in Data Mining: Concepts and Techniques, 3rd ed., The Morgan Kaufmann Series in Data Management Systems, San Francisco, CA: Morgan Kaufmann, 2012, pp. 443–495. doi: 10.1016/B978-0-12-381479-1.00010-1.

[24] F. M. J. M. Shamrat, Z. Tasnim, I. Mahmud, N. Jahan, and N. I. Nobel, “Application of K-Means clustering algorithm to determine the density of demand of different kinds of jobs,” International Journal of Scientific and Technology Research, vol. 9, no. 2, pp. 2550–2557, Feb. 2020.

[25] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, “An extensive comparative study of cluster validity indices,” Pattern Recognition, vol. 46, no. 1, pp. 243–256, Aug. 2012, doi: 10.1016/j.patcog.2012.07.021.

[26] A. R. Mamat, F. S. Mohamed, M. A. Mohamed, N. M. Rawi, and M. I. Awang, “Silhouette index for determining optimal k-means clustering on images in different color models,” International Journal of Engineering & Technology, vol. 7, no. 2.14, p. 105, Apr. 2018, doi: 10.14419/ijet.v7i2.14.11464.

[27] D. L. Davies and D. W. Bouldin, “A cluster separation measure,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/tpami.1979.4766909.

[28] G. Brock, V. Pihur, S. Datta, and S. Datta, “CLValid: an R package for cluster validation,” DOAJ (DOAJ: Directory of Open Access Journals), Jan. 2008, [Online]. Available: https://doaj.org/article/433bec96caf64699b16c19bb7ba36a53

[29] Y. Fan, S. Guo, W. Dai, C. Chen, C. Zhang, and X. Zheng, “Individual-level socioeconomic status and cataract-induced visual disability among older adults in China: the overview and urban-rural difference,” *Front. Public Health*, vol. 12, Art. no. 1289188, Jan. 2024. doi: 10.3389/fpubh.2024.1289188.