BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1685–1694
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1685-1694

A TWO-STEP CLUSTER FOR CLASSIFYING PROVINCES IN INDONESIA BASED ON ENVIRONMENTAL QUALITY

Umi Mahmudah [1*], Muhammad Safiih Lola[2]

[1]*Faculty of Tarbiyah and Educational Sciences, UIN K.H. Abdurrahman Wahid Pekalongan Pekalongan, Central Java, Indonesia*

[2]*Mathematics Sciences, Faculty of Ocean Engineering Technology and Informatics, Malaysia University, Terengganu, Malaysia*

Corresponding author's e-mail: * **umi.mahmudah@uingusdur.ac.id**

Article History:

Received: 11th May 2023
Revised: 11th August 2023
Accepted: 21st August 2023

Keywords:

Clustering;
Two-step methods;
Environmental quality

ABSTRACT

The main objective of this study was to conduct a cluster analysis of the environmental health index in Indonesia for all the provinces. Clustering the environmental health index was important to reveal regional disparities, target and intervention policies, monitor progress over time, and allocate resources more effectively for improved environmental health outcomes. In this study, a sample of 34 units was utilized, encompassing all provinces in Indonesia. The environmental health index was clustered based on five indicators, namely Water Quality Index, Air Quality Index, Soil Quality Index, Marine Quality Index, and Land Cover Quality Index. This research used the two-stage clustering method, which was unique in combining both hierarchical and non-hierarchical clustering methods to produce a more accurate and reliable solution. Four clusters were determined to group provinces in Indonesia based on the environmental health index. The analysis found that the quality of clustering was in the fair but close to good category. The clustering results showed that 32% of the provinces were in cluster 4 and 26.5% of the provinces were in cluster 1. Then, 23.5% and 17.6% of the provinces were in clusters 2 and 3, respectively. In addition, two indicators were found to be the most predictive of the overall clustering solution, namely the Soil Quality Index and the Land Cover Quality Index. The results also implied that provinces in cluster 3 had the lowest environmental quality so they must improve it by looking at provinces in cluster 4, which was the group with the best environmental quality index

How to cite this article:

U. Mahmudah and M. S. Lola., "A TWO-STEP CLUSTER FOR CLASSIFYING PROVINCES IN INDONESIA BASED ON ENVIRONMENTAL QUALITY," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1685-1694, September, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Indonesia has faced many environmental challenges such as deforestation, air and water pollution, waste management issues, and others [1]. These factors have contributed to a decline in the country's environmental health index and have had negative impacts on the health and well-being of its citizens. These environmental challenges have had a range of negative impacts on the health and well-being of Indonesian citizens, and on the country's economy and environment. Efforts are being made by the Indonesian government and various organizations to address these issues [2], [3], but much work remains to be done to improve the environmental health index in Indonesia.

According to Statistics Indonesia, the environmental quality index in 2021 increased by 1.18 points due to an increase in the air quality index and the seawater quality index. Twenty-seven provinces met the 2021 index score target, while seven provinces did not meet the target [4]. However, according to the achievement of the environmental quality index per province in 2021, there were 22 provinces with good predicates in 2020, then increased to 24 provinces in 2021. Meanwhile, there are 10 provinces with a moderate predicate from 12 provinces in the previous year. DKI Jakarta, West Java, and Banten are the three provinces with the lowest environmental quality index, at 64.14, 65.66, and 66.27 respectively. The Environmental Quality Index score has shown an increasing trend over the past three years. The index value in 2021 is 71.45, which falls into the "good" category [4].

The main objective of this research is to cluster each province in Indonesia based on the environmental quality index. Cluster analysis is a common method used in environmental and health research, as it allows for the grouping of data into meaningful clusters based on similarities and differences. Clustering the environmental health index by province can reveal regional disparities in environmental health and highlight areas that need more attention and resources. Besides, it can help decision-makers target policies and interventions in regions where they are most needed, making them more effective and efficient. Furthermore, it allows for monitoring of changes and progresses over time, and measuring the impact of policies and interventions. By understanding the regional differences in environmental health, resources can be allocated more effectively, leading to improved environmental health outcomes for communities.

Therefore, it is crucial to conduct a cluster analysis for the environmental quality index in Indonesia based on each province because this information can be critical for decision-makers and policymakers in determining where to allocate resources and prioritize efforts to improve environmental health. Unfortunately, not many previous studies have been found on this theme. However, previous studies have utilized cluster analysis to assess the environmental quality in Indonesia, but have not specifically classified the environmental quality index. A study used cluster analysis of the groundwater data to determine the quality of the water [5]. A study was conducted to assess the water quality trends in Jakarta as well as to perform a clustering analysis. The results of the cluster analysis suggested three groups for biochemical oxygen demand (BOD) and total suspended solids (TSS), and four groups for dissolved oxygen (DO) [6].

However, the use of two-step clustering method has not been done in the classification of environmental quality index in Indonesia. In fact, this method is believed to be able to provide results with good accuracy. Therefore, this research is important to provide additional insights into Indonesia's environmental quality. By grouping areas with similar environmental health indices into clusters, this analysis provides a more nuanced understanding of the environmental health situation in a country, highlighting areas that may require additional attention and resources.

2. RESEARCH METHODS

2.1 Data and Variables

This study used environmental health index data for each province in Indonesia published by the Indonesian Central Bureau of Statistics in 2021. There were five indicators, namely Water Quality Index (WQI), Air Quality Index (AQI), Soil Quality Index (SQI), Marine Quality Index (MQI), and Land Cover Quality Index (LCQI). Water Quality Index (WQI) is a composite index that evaluates the overall water quality of a particular area based on several parameters. It is often used as a part of the environmental health index to assess the quality of water sources in a region, such as rivers, lakes, or groundwater. WQI is calculated by considering multiple water quality parameters, including pH, temperature, dissolved oxygen, conductivity, and the presence of pollutants, among others. These parameters are assigned scores.

The Air Quality Index (AQI) is a numerical index that provides information about the quality of air in a specific area. It is used to measure the level of air pollution and its potential impacts on human health and the environment. The AQI is calculated based on the concentrations of major air pollutants, including particulate matter, ozone, nitrogen dioxide, sulfur dioxide, and carbon monoxide. The AQI is used to indicate the level of air pollution, with a higher AQI value indicating worse air quality. The Soil Quality Index (SQI) is a measure of the overall soil quality and its ability to support plant growth, maintain water quality, and store carbon. It is a component of the Environmental Quality Index (EQI), which assesses the overall health and sustainability of the environment. SQI is determined by evaluating various physical, chemical, and biological properties of soil, such as nutrient content, organic matter, structure, and soil biota.

The Marine Quality Index (MQI) is a measure of the overall health of marine ecosystems and their ability to provide essential ecosystem services, such as food production, climate regulation, and biodiversity support. It is a component of the Environmental Quality Index (EQI), which assesses the overall health and sustainability of the environment. MQI is determined by evaluating various physical, chemical, and biological parameters of marine ecosystems, such as water quality, primary productivity, and biodiversity. The Land Cover Quality Index (LCQI) is a measure of the condition of land covers, including forests, croplands, grasslands, and urban areas, and its ability to provide important ecosystem services. It is a component of the Environmental Quality Index (EQI), which assesses the overall health and sustainability of the environment. LCQI is determined by evaluating the extent, condition, and diversity of different land cover types, as well as their ability to provide essential ecosystem services, such as carbon sequestration, water regulation, and biodiversity support. **Table 1** below shows an overview of the environmental health index based on the 5 indicators above.

Table 1. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Water Quality Index	34	43,09	58,37	53,26	4,20
Air Quality Index	34	66,52	95,60	88,45	5,63
Soil Quality Index	34	26,25	100,00	61,35	20,79
Marine Quality Index	34	70,34	87,55	81,64	4,24
Land Cover Quality Index	34	26,25	100,00	61,51	20,71

Table 1 above illustrates an overall depiction of the environmental health index based on the five utilized indicators in Indonesia, wherein the lowest mean value is attributed to the Water Quality Index (WQI) at 53.26, whereas the highest mean value is associated with the Air Quality Index (AQI) at 88.45. This signifies that among the five indicators of the environmental health index, the air quality index exhibits the most favorable performance compared to the other indicators.

2.2 Two-step cluster

Two-step cluster analysis is a data analysis technique that is used to group similar observations into clusters based on the similarities and differences among the variables **[7]**, **[8]**. This method is commonly used in fields such as sociology, psychology, and market research to identify patterns and relationships among large datasets. The two-step approach is unique in that it combines both a hierarchical and a non-hierarchical clustering method to produce a more accurate and reliable solution **[9]**, **[10]**. One advantage of two-step cluster analysis is that it can handle complex datasets with many variables and can produce more accurate results than traditional clustering methods. The hierarchical method is useful for exploring the structure of the data, while the non-hierarchical method can be used to refine the results and produce a final clustering solution.

The first step of two-step cluster analysis is a hierarchical clustering method, where a *dendrogram* is created to visualize the relationships among the observations. This *dendrogram* is used to determine the number of clusters that should be formed and the composition of the clusters. The second step is a non-hierarchical clustering method, such as k-means clustering, where the observations are assigned to the clusters based on the similarities among the variables. This research used the following stages in classifying Indonesian provinces based on environmental quality using the two-step cluster technique **[11]**:

Stage 1: pre clustering

In this stage, the Cluster Feature (CF) Tree has been created, made up of nodes and branches that have leaf entries. The sub-clusters are represented by these leaf entries, which are evaluated when a new input object is encountered. This is done by measuring the distance between the new entry and the existing leaf entries. If the distance is close, the object is added to the first-built leaf entry. If the distance is far, a new leaf entry is created for the object. This process involves examining each data vector individually and determining whether it should be added to an existing leaf entry or if a new one should be created. When the capacity of a branch in the CF Tree has been reached, the node is divided into two based on the furthest object within it. The remaining objects are allocated to either node based on their proximity to each. This process continues until all the objects have been organized into clusters.

Stage 2: Number of Clusters Optimization

The Bayesian Information Criterion (BIC) can be used to determine the maximum number of clusters. The formula for BIC is provided as follows:

$$BIC(j) = -2\sum_{i=1}^{j} \xi_i + m_j \log N \tag{1}$$

Where

$$m_j = j\left\{2K^A + \sum_{K=1}^{K^B}(L_k - 1)\right\} \tag{2}$$

N was the number of the observation while j represented the cluster number, K^A and K^B were the numbers of numerical and categorical variables, respectively. L_k was the category number of the k-th categorical variable.

The optimal number of clusters can be calculated by determining the ratio of the distance measure using the following formula:

$$R(j) = \frac{k_{j-1}}{k_j} \tag{3}$$

Where $k_j = l_{j-1} - l_j$. This was the distance when j clusters are combined into j-1 clusters.

And

$$l_v = \frac{m_v \log(N) - BIC(v)}{2} \tag{4}$$

Where $v = j, j-1$.

Figure 1 is the flowchart for the procedure of the two-step cluster technique:

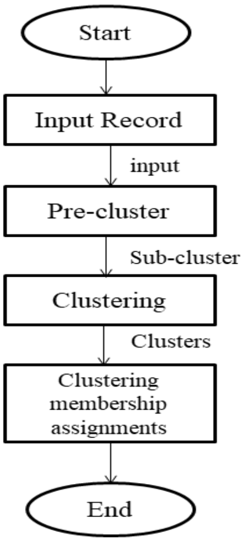

Figure 1. Flowchart of two-step cluster

2.3 Log-Likelihood Distance

The formula for log-likelihood distance to measure the similarity between objects because of its ability to measure the similarity between mixed categorical and numerical variables can be expressed as follows [12]:

$$d(j,s) = \xi_j + \xi_s - \xi_{<j,s>} \tag{5}$$

Where $d(j,s)$ was the distance of cluster j and s while $\xi_{<j,s>}$ was the marginal index for cluster j and s. Thus, from the Equation (5), the log-likelihood for cluster v can be written as follows:

$$\xi_v = N_v \begin{bmatrix} \sum_{k=1}^{K^A} \frac{\log\left(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2\right)}{2} \\ \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log\left(\frac{N_{vkl}}{N_v}\right) \end{bmatrix};\ \begin{matrix} v = j,s \\ < j,s > \end{matrix} \tag{6}$$

N_v was the number of objects in cluster v, $\hat{\sigma}_k^2$ was the variance predictor for the k-th numerical variable for all objects. N_{vkl} was the number of objects in clusters v for k-th numerical variable and l-th categorical variable.

3. RESULTS AND DISCUSSION

Table 2 below showed the cluster profile based on centroids. A centroid was a representative point or mean of a group of points in a cluster. It was used to summarize the characteristics of a cluster and to define its location in the feature space. The process of finding the centroid of a cluster involved calculating the average of the feature values of all the points in the cluster. Centroids played an important role in many clustering algorithms, such as k-means and k-medoids, as they were used to determine the optimal grouping of data points into clusters.

Table 2. Centroids

		WQI		AQI		SQI		MQI		LCQI	
		Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Cluster	1	52.11	3.39	90.23	0.88	60.55	8.07	79.38	3.24	60.79	7.88
	2	56.38	2.156	87.97	2.67	45.91	8.33	82.34	3.61	46.18	8.19
	3	47.48	4.27	80.04	8.50	38.15	9.06	83.113	4.23	38.29	9.33
	4	55.08	1.97	91.91	2.07	85.89	10.38	82.16	5.12	85.92	10.37
	Combined	53.26	4.20	88.45	5.63	61.35	20.79	81.64	4.24	61.51	20.70

The process of finding the centroid of a cluster involved calculating the average of the feature values of all the points in the cluster. Centroids played an important role in many clustering algorithms, such as k-means and k-medoids, as they were used to determine the optimal grouping of data points into clusters. Table 1 showed that four clusters emerged from the provincial classification based on the Environmental Quality Index. The four clusters were based on the similarities and differences in each province's characteristics as revealed by the five indicators used: Water Quality Index, Air Quality Index, Soil Quality Index, Marine Quality Index, and Land Cover Quality Index.

The centroids showed that the clusters were well separated by the continuous variables. Provinces in cluster 1 were the worst in managing environmental quality. Provinces in cluster 2 and cluster 3 were moderately managed. Nevertheless, the differences and similarities in characteristics can be seen between the two clusters. Provinces in cluster 4 were the best in managing environmental quality, in which they had the best score of the environmental quality index. Furthermore, the analysis revealed the majority of provinces were in cluster 4, which was 32% or 11 provinces. A total of 26% or 9 provinces were in Cluster 1. Then cluster 2 was composed of 8 provinces or 24%, and cluster 3 was the cluster with the smallest number of members, namely 6 provinces or 18%.

The analysis using two-step clustering revealed that the environmental quality index of Indonesian provinces was divided into four clusters. Provinces in cluster 1 were Riau, Riau Islands, Bengkulu, West Sumatra, West Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, and North

Sulawesi. Cluster 2 consisted of North Sumatra, South Sumatra, Bangka Belitung, Lampung, East Java, Bali, East Nusa Tenggara, and South Sulawesi. Cluster 3 covered Jambi, Banten, DKI Jakarta, West Java, Central Java, and the Special Region of Yogyakarta. And cluster 4 consisted of Aceh, East Kalimantan, North Kalimantan, Southeast Sulawesi, Central Sulawesi, West Sulawesi, Gorontalo, Maluku, North Maluku, Papua, and West Papua.

Based on an in-depth analysis of the Environmental Quality Index published by Statistics Indonesia, it was found that Cluster 4 was a group of provinces with the highest Environmental Quality Index, where all the members of the cluster had an index category in the good category. The province with the best index score was West Papua province with 81.80 followed by North Kalimantan province with a score of 80.85. It was important to inform they were the two provinces with the best scores in Indonesia. Furthermore, the findings also suggested that cluster 3 was the cluster with members that had the lowest index scores, in which the Jakarta Special Capital Region was the province with the poorest environmental quality, with a score of 54.43. Meanwhile, West Java province was the province with the second worst environmental quality, with a score of 62.68. These were also the two provinces with the lowest scores in Indonesia.

Since these scores did not merely provide a ranking of environmental quality, but can also be an indication of efforts to improve environmental quality in the provincial and national areas, the above findings were crucial because they provided a very clear picture of the provinces that required serious attention from the government to improve and enhance environmental quality for each province. Likewise, provinces in cluster 3 can conduct benchmarking in the context of comparative studies with provinces in cluster 4 related to environmental quality management.

The results of the cluster analysis also found that the quality of the classification of provinces in Indonesia based on the environmental quality index into four clusters is in the "Fair but close to good" category, where the average silhouette was 0.4. It implied that the clustering of data points was very close to a level being considered "Good" but there may still be room for improvement. The results also revealed that the ratio of size based on the largest cluster to the smallest cluster is about 1.83. This referred to the relationship between the size of the largest and smallest clusters in the clustering solution. The size ratio was calculated as the ratio of the number of data points in the largest cluster to the number of data points in the smallest cluster. The result provided a number of 1.83 which was less than 2, which was the criterion for acceptability.

The results also revealed that the ratio of size based on the largest cluster to the smallest cluster is about 1.83. This referred to the relationship between the size of the largest and smallest clusters in the clustering solution. The size ratio was calculated as the ratio of the number of data points in the largest cluster to the number of data points in the smallest cluster. The result provided a number of 1.83 which was less than 2, which was the criterion for acceptability. The number indicated that the largest cluster was approximately 1.83 times larger than the smallest cluster. It was important to note that this information could be used to assess the balance of the clustering solution and determine if the size ratio was acceptable or needs to be adjusted. It can thus be said that the clustering solution was balanced and met the acceptance criterion, which required no further adjustments.

Figure 2 below showed the detailed information for each cluster based on the input variables used, namely the five indicators of the environmental quality index in Indonesia. Input variables here referred to the data used as input in the clustering process. In two-step clustering, this input was received by the clustering algorithm and used to create a cluster of data **[8]**, **[13]**. The input was usually a numerical vector dataset that represents the features of the data to be clustered.

Soil Quality Index 60.55	Soil Quality Index 45.91	Soil Quality Index 38.15	Soil Quality Index 85.89
Land Cover Quality Index	Land Cover Quality Index	Land Cover Quality Index	Land Cover Quality Index
Water Quality Index 52.11	Water Quality Index 56.38	Water Quality Index 47.48	Water Quality Index 55.08
Air Quality Index 90.23	Air Quality Index 87.98	Air Quality Index 80.04	Air Quality Index 91.91
Marine Quality Index 79.38	Marine Quality Index 82.34	Marine Quality Index 83.11	Marine Quality Index 82.16

Figure 2. Information for Each Cluster

Some important information from the results of the two-step cluster analysis is shown in **Figure 2** above. The darkest blue in the first two rows of the input section indicated the two most predictive input variables. Note that the darker the color, the higher the predicted value. Conversely, lighter colors indicated lower predictor importance values. In the analysis, it was found that the soil quality index and the land cover quality index had the same predictor importance score of 1.00. Then, it was also found that the predictive importance score for the water quality index and the air quality index also gave the same score, which amounted to 0.45, while the lowest predictive importance score was for the marine quality index, which amounted to 0.05.

Furthermore, **Figure 2** also showed the mean values of the different variables from the cluster analysis. For example, in cluster 1, it was found that the mean soil quality index was 60.55, while the mean land cover quality index gave a value of 60.79. Then, the average of the water quality index and the air quality index gave an average of 52.11 and 90.23, respectively, while the marine quality index had an average of 79.38. In cluster 2, the mean values of the input variables were found to be as follows: soil quality index and land cover quality index of 45.91 and 46.18, respectively. Then, the average values of the water quality index and air quality index were 56.38 and 87.98, respectively, and the average marine quality index was 82.34.

The results also revealed that two indicators were the most important predictors of the overall clustering solution, namely the index of the quality of the soil and the index of the quality of the land cover. This result could be due to direct indicators of the health of the soil and the land, which in turn, had a significant impact on the overall health of the environment. Soil and land quality played a crucial role in maintaining ecosystem stability and preserving biodiversity, as well as providing essential ecosystem services such as food production and water filtration **[14]**. The Soil Quality Index and Land Cover Quality Index provided a comprehensive measure of the health of these important components of the environment, making them valuable indicators for determining the overall environmental quality index **[15]**. In addition, the Soil Quality Index and Land Cover Quality Index were found to be the most predictive of the overall clustering of the environmental health index due to their relevance, data availability, indicator diversity, and ease of measurement.

In the context of a two-step cluster method, predictor importance referred to the measurement of the contribution of each predictor or variable to the overall clustering solution **[16]**. This can help to identify the most important predictors for explaining the patterns and structure in the data and make the clustering solution more interpretable. Assessing predictor importance can also help to identify and remove redundant or irrelevant variables, improving the efficiency and interpretability of the clustering solution.

The following figure showed the comparison of the four input variables in each cluster.

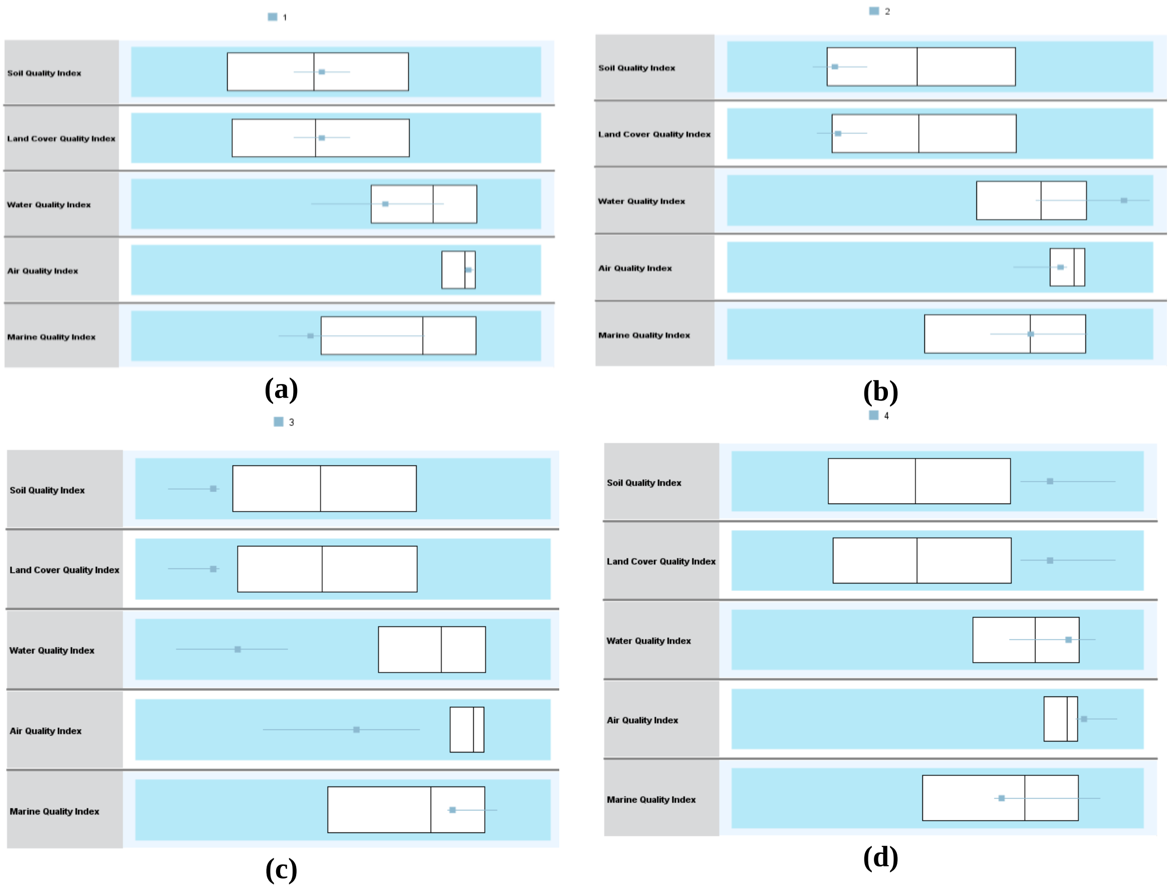

Figure 3. Cluster Comparison
(a) Cluster 1, (b) Cluster 2, (c) Cluster 3, (d) Cluster 4

The figure above suggested another important finding related to two-step clustering, namely the median value, where its value in cluster 1 for each input variable was as follows: soil quality index = 60.54; land cover quality index = 60.54; water quality index = 52.56; air quality index = 90.44; marine quality index = 77.87. Then, the median in cluster 2 for each input variable was as follows: soil quality index = 44.87; land cover quality index = 45.42; water quality index = 57.32; air quality index = 89.28; marine quality index = 82.60. Furthermore, the analysis also found that the median value in cluster 3 for each input variable is as follows: soil quality index = 40.08; land cover quality index = 40.08; water quality index = 46.85; air quality index = 82.01; marine quality index = 83.48. Meanwhile, the median in cluster 4 for each input variable was as follows: soil quality index = 83.22; land cover quality index = 83.22; water quality index = 55.58; air quality index = 91.38; marine quality index = 81.61.

In the context of a two-step cluster method, the median can refer to the median of a variable or set of variables used for dimensionality reduction [10]. The median can be used as a central tendency measure for transforming the data, for example by transforming each variable into its deviations from the median. This can help to reduce the effects of outliers or extreme values and make the data easier to cluster. The median can also be used as a cutting point to divide the data into two groups, which can then be further analyzed or processed in the next step of the clustering method.

Broadly speaking, to have improved environmental quality in Indonesia, the provinces that belonged to Cluster 3 could have learned from those in Cluster 4, which was the best-performing group in terms of environmental quality. They could have undertaken a detailed analysis of the factors that contributed to the superior environmental quality of Cluster 4. This analysis could have involved identifying the specific environmental indicators and practices that were associated with Cluster 4 and comparing them with those of Cluster 3. By doing so, provinces belonging to Cluster 3 could have identified areas for improvement and implemented changes to their policies and practices to enhance their environmental quality. Additionally, provinces in Cluster 3 could have sought out guidance and support from experts and organizations with expertise in environmental management and sustainability to help them improve their practices and achieve better environmental outcomes. The goodness of the cluster model was presented by a diminished level of within-cluster variance serves as an indicative hallmark of strong compactness within the clusters. This signifies that the data points grouped within each cluster exhibit a heightened degree of similarity and proximity to one another concerning the designated clustering criteria. In essence, this reduced within-cluster variation underscores the efficacy of the clustering algorithm in encapsulating homogeneous data points within cohesive and distinct clusters, thereby contributing to the overall robustness of the clustering solution.

4. CONCLUSIONS

Clustering the environmental health of Indonesia by provinces provided valuable insights into the state of the environment, informed policy-making, and helped ensure that resources were used effectively. In conclusion, cluster analysis of a country's environmental health index was an important tool for evaluating the state of the environment and protecting public health. By highlighting areas that required additional attention and resources, and by providing insight into the underlying factors that contributed to differences in environmental health, cluster analysis helped to guide efforts to improve the environment and protect public health. The results of this study provided important findings, where provinces in cluster 4 can be used as a model for other provinces in maintaining good environmental quality. They were Aceh, East Kalimantan, North Kalimantan, Southeast Sulawesi, Central Sulawesi, West Sulawesi, Gorontalo, Maluku, North Maluku, Papua, and West Papua. Meanwhile, provinces in cluster 3 required very serious attention from the Indonesian government so that environmental quality can be improved. Cluster 3, on the other hand, needed very serious attention from the Indonesian government for improving the quality, which was positively related to the health of its people.

REFERENCES

[1] S. Nomani, M. Rasel, and I. K. Reedoy, "Industrial Development and Climate Change: A Case Study of Bangladesh," *Indones. J. Innov. Appl. Sci.*, vol. 2, no. 1, pp. 68–79, 2022, doi: https://doi.org/10.47540/ijias.v2i1.428.

[2] S. Partelow and K. Nelson, "Social networks, collective action and the evolution of governance for sustainable tourism on the Gili Islands, Indonesia," *Mar. Policy*, vol. 112, 2020, doi: https://doi.org/10.1016/j.marpol.2018.08.004.

[3] J. Koehrsen, "Muslims and climate change: How Islam, Muslim organizations, and religious leaders influence climate change perceptions and mitigation activities," *Wiley Interdiscip. Rev. Clim. Chang.*, vol. 12, no. 3, p. e702, 2021, doi: https://doi.org/10.1002/wcc.702.

[4] Ministry of Environment and Forestry, "Laporan Kinerja Lingkungan Hidup (Environmental Performance Report)," 2021.

[5] T. R. N. Amanah, T. T. Putranto, and M. Helmi, "Application of cluster analysis and principal component analysis for assessment of groundwater quality—A study in Semarang, Central Java, Indonesia," in *IOP conference series: earth and environmental science*, 2019, vol. 248, no. 1, p. 12063, doi: 10.1088/1755-1315/248/1/012063.

[6] P. Luo *et al.*, "Water quality trend assessment in Jakarta: A rapidly growing Asian megacity," *PLoS One*, vol. 14, no. 7, p. e0219009, 2019, doi: https://doi.org/10.1371/journal.pone.0219009.

[7] M. Benassi *et al.*, "Using two-step cluster analysis and latent class cluster analysis to classify the cognitive heterogeneity of cross-diagnostic psychiatric inpatients," *Front. Psychol.*, vol. 11, p. 1085, 2020, doi: https://doi.org/10.3389/fpsyg.2020.01085.

[8] V. Lytvynenko, I. Lurie, J. Krejci, M. Voronenko, N. Savina, and M. A. Taif, "Two Step Density-Based Object-Inductive Clustering Algorithm.," in *MoMLeT*, 2019, pp. 117–135.

[9] A. Iodice D'Enza and F. Palumbo, "Iterative factor clustering of binary data," *Comput. Stat.*, vol. 28, pp. 789–807, 2013, doi: https://doi.org/10.1007/s00180-012-0329-x.

[10] M. Van de Velden, A. Iodice D'Enza, and A. Markos, "Distance-based clustering of mixed data," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 11, no. 3, p. e1456, 2019, doi: https://doi.org/10.1002/wics.1456.

[11] A. Supandi, A. Saefuddin, and I. D. Sulvianti, "Two step cluster application to classify villages in Kabupaten Madiun based on village potential data," *Xplore J. Stat.*, vol. 10, no. 1, pp. 12–26, 2021.

[12] T. Chiu, D. Fang, J. Chen, Y. Wang, and C. Jeris, "A robust and scalable clustering algorithm for mixed type attributes in large database environment," in *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, 2001, pp. 263–268, doi: https://doi.org/10.1145/502512.502549.

[13] I. N. Martínez, J. M. Morán, and F. J. Peña, "Two-step cluster procedure after principal component analysis identifies sperm subpopulations in canine ejaculates and its relation to cryoresistance," *J. Androl.*, vol. 27, no. 4, pp. 596–603, 2006, doi: https://doi.org/10.2164/jandrol.05153.

[14] R. Haines-Young and M. Potschin, "The links between biodiversity, ecosystem services and human well-being," *Ecosyst. Ecol. a new Synth.*, vol. 1, pp. 110–139, 2010.

[15] J. B. Kowalska, R. Mazurek, M. Gąsiorek, and T. Zaleski, "Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination–A review," *Environ. Geochem. Health*, vol. 40, pp. 2395–2420, 2018, doi: https://doi.org/10.1007/s10653-018-0106-z.

[16] A. Tkaczynski, "Segmentation using two-step cluster analysis," in *Segmentation in social marketing*, Springer, 2017, pp. 109–125.