Available online at
www.heca-analitika.com/ljes

Leuser Journal of Environmental Studies

Vol. 2, No. 1, 2024

Environmental and Economic Clustering of Indonesian Provinces: Insights from K-Means Analysis

Teuku Rizky Noviandy [1], Irsan Hardi [2], Zahriah Zahriah [3], Rahmi Sofyan [4], Novi Reandy Sasmita [5], Iin Shabrina Hilal [6] and Ghalieb Mutig Idroes [7,*]

[1] Interdisciplinary Innovation Research Unit, Graha Primera Saintifika, Aceh Besar 23771, Indonesia; trizkynoviandy@gmail.com (T.R.N.)
[2] Economic Modeling and Data Analytics Unit, Graha Primera Saintifika, Aceh Besar 23371, Indonesia; irsan.hardi@gmail.com (I.H.)
[3] Department of Architecture and Urban Planning, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; zahriah@usk.ac.id (Z.Z.)
[4] Department of Early Childhood Education, Faculty of Teacher Training and Education, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rahmisofyan@usk.ac.id (R.S.)
[5] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah, Banda Aceh 23111, Indonesia; novireandys@usk.ac.id (N.R.S.)
[6] Department of Civil Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; shabrinahilal10@gmail.com (I.S.H.)
[7] Energy and Green Economics Unit, Graha Primera Saintifika, Aceh Besar 23371, Indonesia; ghaliebidroes@outlook.com (G.M.I.)

* Correspondence: ghaliebidroes@outlook.com

Article History

Received 21 February 2024
Revised 13 April 2024
Accepted 22 April 2024
Available Online 29 April 2024

Keywords:
Clustering
AQI
GRDP
Electricity

Abstract

Indonesia's archipelago presents a distinctive opportunity for targeted sustainable development due to its complex interplay of economic advancement and environmental challenges. To better understand this dynamic and identify potential areas for focused intervention, this study applied K-means clustering to 2022 data on the Air Quality Index (AQI), electricity consumption, and Gross Regional Domestic Product (GRDP). The analysis aimed to delineate the provinces into three distinct clusters, providing a clearer picture of the varying levels of economic development and environmental impact across the nation's diverse islands. Each cluster reflects specific environmental and economic dynamics, suggesting tailored policy interventions. The results show that for provinces in Cluster 1, which exhibit moderate environmental quality and lower economic activity, the introduction of sustainable agricultural enhancements, eco-tourism, and renewable energy initiatives is recommended. Cluster 2, marked by higher economic outputs and moderate environmental conditions, would benefit from the implementation of smart urban planning, stricter environmental controls, and the adoption of clean technologies. Finally, Cluster 3, which includes highly urbanized areas with robust economic growth, requires expanded green infrastructure, improved sustainable urban practices, and enhanced public transportation systems. These recommendations aim to foster balanced economic growth while preserving environmental integrity across Indonesia's diverse landscapes.

1. Introduction

As the largest archipelagic country in the world, Indonesia provides a fascinating case study for understanding regional economic disparities, sustainability challenges, and the impacts of economic policies [1–3]. With over 17,000 islands spanning a vast area in Southeast Asia, Indonesia showcases a complex landscape where economic progress intersects with environmental concerns [4, 5]. As the nation strives to balance economic growth with environmental protection, analyzing its development path offers valuable insights into navigating the path toward sustainable development [6–8].

In examining Indonesia's intricate socio-economic and environmental landscape, a comprehensive understanding can be achieved by exploring into key factors such as the Air Quality Index (AQI), Gross Regional Domestic Product (GRDP), and electricity consumption [9]. These metrics offer valuable insights into the interplay between economic development, environmental sustainability, and energy utilization within the Indonesian provinces. Uncovering the correlations and dependencies among AQI, GRDP, and electricity consumption, we can unveil the complex relationships shaping Indonesia's regional development trajectory and inform strategies for achieving sustainable growth.

The AQI is an essential environmental indicator in this study, providing a quantitative measure of air pollution levels across Indonesia's provinces [10, 11]. The rapid pace of urbanization and the proliferation of industrial activities have exerted a profound influence on AQI values, resulting in striking variations in air quality between industrialized regions and rural areas [12]. By incorporating AQI into the clustering analysis, this study endeavors to uncover the underlying patterns that link specific economic activities and urbanization rates with air pollution levels [13, 14]. The analysis provides valuable insights that assist policymakers in crafting precise environmental regulations and interventions, promotes sustainable urban development and enhances public health policies.

Moreover, electricity consumption emerges as a key factor in examining regional disparities and development levels [15, 16]. Electricity usage not only serves as a proxy for the degree of modernization and industrial capacity of a province but also bears significant implications for the environmental footprint through the choice of energy production sources [17, 18]. By integrating electricity usage data into the clustering algorithm, this research aims to delineate provinces that may require infrastructural enhancements or transitions towards renewable energy sources [15, 18]. The insights derived from this analysis will inform the formulation of energy policies that are not only efficient but also sustainable, aligning with Indonesia's long-term objectives of reducing carbon emissions and promoting energy security in an environmentally conscious manner [16, 19].

Shifting the focus to economic indicators, the GRDP serves as a barometer of the economic vitality and productivity of each province in Indonesia [20–22]. The observed variations in GRDP underscore the unequal distribution of economic resources and development opportunities across the archipelago [23–25]. By clustering provinces based on GRDP, in conjunction with AQI and electricity consumption, this study illuminates the economic growth patterns that align with environmental and electricity consumption dynamics [19]. This comprehensive perspective facilitates a deeper understanding of the complex interdependencies between economic growth and regional sustainability challenges, informing the development of more balanced economic policies that promote equitable growth across all regions.

In analyzing Indonesia's complex socio-economic and environmental data, K-means clustering analysis emerged as a promising method. K-means is an unsupervised machine learning algorithm, which means it autonomously partitions data points into distinct clusters based on similarities in features [26, 27]. Unlike another method, such as hierarchical clustering analysis (HCA), K-means offers advantages in scalability and computational efficiency [28]. K-means reveals inherent patterns and structures within the data, enabling researchers to identify meaningful relationships among socio-economic and environmental factors across Indonesia's provinces, allowing policymakers to tailor strategies and allocate resources effectively according to each region's unique needs and challenges. Additionally, uncovering underlying trends through clustering can inform targeted approaches for promoting sustainable development, mitigating environmental issues, and driving inclusive economic growth across Indonesia's diverse regions [19, 29].

This research aims to investigate the regional disparities and sustainability challenges in Indonesia by employing a k-means clustering approach to analyze key variables, including AQI, GRDP, and electricity consumption. This study contributes significantly to the existing body of knowledge by offering a comprehensive viewpoint on the intricate interplay between economic development, environmental conservation, and energy usage across Indonesia's diverse provinces. The insights generated from this analysis have the potential to inform evidence-

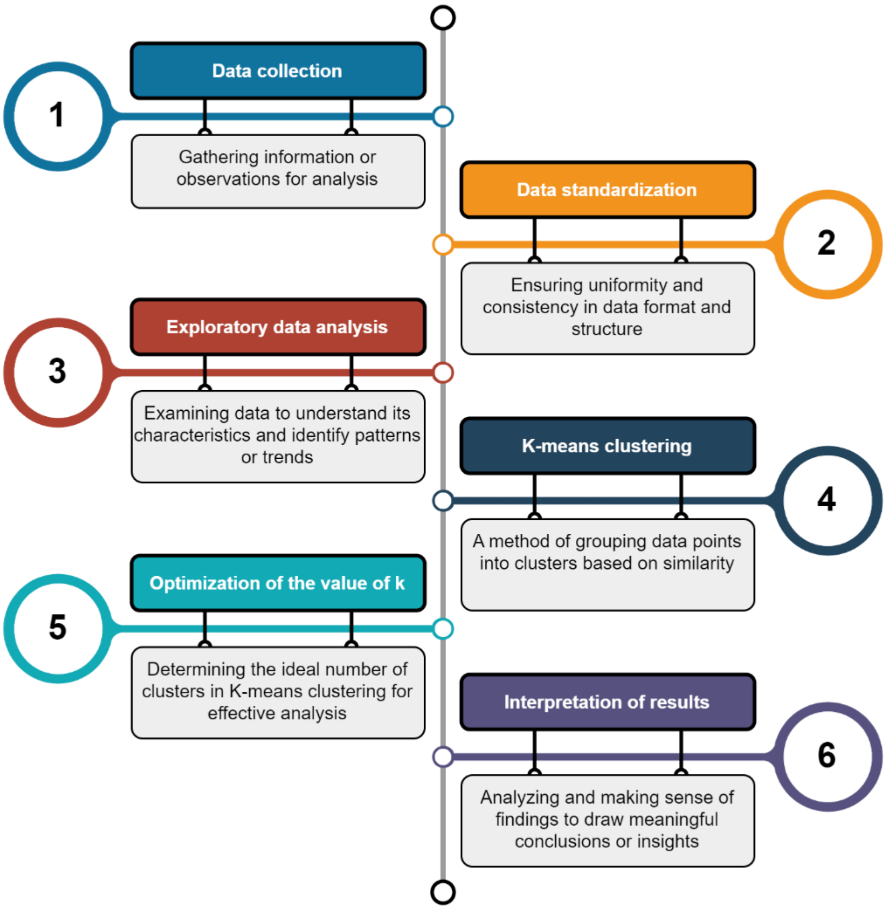

Figure 1. Workflow of this study.

based policymaking and guide the implementation of targeted interventions that promote sustainable growth and equitable development across Indonesia's vast archipelagic landscape.

2. Materials and Methods

The study's workflow, depicted in Figure 1, consists of six well-defined steps. The process begins with the collection of relevant data for analysis. The collected data then undergoes standardization to ensure consistency and facilitate accurate analysis. Once standardized, the data is subjected to an exploratory analysis to uncover patterns and trends that may provide valuable insights. Building upon these insights, the K-means clustering algorithm is applied to group similar data points into meaningful clusters. To ensure optimal clustering results, the number of clusters were carefully tuned. Finally, the results are thoroughly interpreted to draw conclusions, gain actionable insights, and enable informed decision-making based on the findings.

2.1. Data

In this study, we utilize data from 2022 to capture an overview of the AQI, electricity consumption, and GDRP. We chose 2022 data because it offers a more representative dataset after the COVID-19 pandemic, which started in early 2020 and notably affected Indonesia's economy and environmental factors. By focusing on 2022, we ensure that the data reflects a more stable period as the economic sector has begun to recover. The AQI data was collected from the Ministry of Environment and Forestry (KLHK), while the electricity consumption and GRDP data were obtained from the Central Bureau of Statistics (BPS). A synopsis of the data is presented in Table 1.

2.2. K-Means Clustering Analysis

K-means clustering is a popular unsupervised machine learning algorithm used for partitioning a dataset into k distinct, non-overlapping clusters. The algorithm aims to minimize the within-cluster sum of squares (WCSS), which represents the sum of squared distances between each data point and its centroid within the cluster. The K-means algorithm iteratively assigns each data point to the nearest centroid and then recalculates the centroids based on the mean of all data points assigned to each cluster. The objective function of K-means can be defined as shown in Equation 1 [22, 30].

Table 1. Synopsis of the variables.

Variable	Description	Measure	Source
AQI	Serves as a standardized measure to objectively assess and communicate the level of air pollution in a given area, enabling individuals and authorities to understand potential health risks associated with breathing the air.	Index (scale 1-100)	KLHK
Electricity	Serves as a quantifiable measure of energy flow, providing an objective indicator of power generation, distribution, and consumption within a designated area or system.	Gigawatt hour (GWh)	BPS
GRDP	Quantifies the economic output generated within a specific region over a defined period, providing an objective measure of the region's economic activity and performance.	Constant 2010	BPS

$$Minimize: \sum_{i=10}^{k} \sum\nolimits_{x \in C_i} \parallel x - \mu_i \parallel^2 \qquad (1)$$

Where k is the number of clusters, C_i is the set of data points assigned to cluster i, and μ_i is the centroid of cluster i.

In this study, we utilize Python version 3.10.9 and scikit-learn version 1.2.0 [31] to perform K-means clustering. We initialize K-means using the K-means++ algorithm, which is a smart initialization technique that selects initial cluster centroids to speed up convergence and improve clustering quality [32]. K-means++ ensures that the initial centroids are spread out, reducing the likelihood of converging to suboptimal solutions. Additionally, we set the maximum number of iterations to 300 to limit the algorithm's runtime and prevent it from running indefinitely. This parameter controls the maximum number of iterations K-means will perform before converging to a solution.

2.3. Data Preprocessing

In this study, we preprocess the data by applying standardization, which involves removing the mean and scaling to unit variance [33]. This step is important because K-Means clustering relies on the concept of minimizing the sum of squared distances between data points and their respective cluster centroids. Since K-Means uses Euclidean distance as a similarity measure, the algorithm is sensitive to the scale of features. Features with larger scales will have a disproportionate impact on the clustering process, potentially leading to inaccurate cluster assignments. Standardizing the features ensures that each feature contributes equally to the computation of distances, thereby improving the effectiveness of the clustering algorithm.

The standardization involves transforming each feature x_i in the dataset to have a mean of 0 and a standard deviation of 1. Mathematically, this transformation can be represented as shown in Equation 2.

$$z_i = \frac{x_i - \mu}{\sigma} \qquad (2)$$

Where z_i is the standardized value of feature x_i, μ is the mean of feature x_i across all data points, and σ is the standard deviation of feature x_i across all data points.

2.4. Determining the Number of Clusters

In this study, one of the important steps is determining the optimal number of clusters (k) for the dataset. We use elbow method for this purpose. It involves plotting the WCSS against the number of clusters (k) ranging from 1 to a predefined maximum value k = 10. The Elbow Method works on the principle that as the number of clusters increases, WCSS tends to decrease because the data points are closer to their respective centroids. However, beyond a certain point, adding more clusters doesn't significantly reduce WCSS, resulting in a less pronounced decrease in the rate of WCSS reduction. This point is termed as the "elbow," indicating the optimal number of clusters. Once the elbow point is identified, it provides a reasonable estimate of the optimal k value.

3. Results and Discussion

3.1. Exploratory Data Analysis

3.1.1. Descriptive Statistics

The descriptive statistics for three key variables, AQI, Electricity, and GRDP, are presented in Table 2. These statistics provide a comprehensive overview of the central tendency, dispersion, and range of the data. The average AQI across the provinces is 89.184, with a relatively narrow standard deviation of 5.556, indicating that the AQI values are closely clustered around the mean. The minimum AQI recorded is 68.06, and the maximum is 95.79, suggesting a moderate range of variation in air quality across provinces. Electricity distribution exhibits a wide range of values, indicated by a large standard deviation of 13196.1 compared to the mean of 8051.81 GWh. This significant variance highlights the disparity in electricity access and consumption across different provinces. The minimum electricity distributed is only 192.11 GWh, whereas the maximum reaches a substantial 56226.1 GWh, pointing to vast differences in infrastructural development or industrial activities. The

Table 2. Descriptive statistics of the data.

Variable	AQI	Electricity	GRDP
Mean	89.184	8051.81	348.240
Std. Dev	5.556	13196.1	496.903
Min	68.06	192.11	30.284
25%	89.167	1230.56	78.713
50%	90.485	3094.43	149.600
75%	91.775	5776.36	356.536
Max	95.79	56226.1	1953.489

GRDP, represented in billions of IDR, also shows a broad dispersion, as evidenced by a standard deviation of 496.903 against a mean of 348.240 billion IDR. The economic output ranges from a low of 30.284 billion IDR to a high of 1953.489 billion IDR, underscoring the economic diversity among the provinces. The quartile values further illustrate this disparity, with 25% of the provinces having a GRDP below 78.713 billion IDR, while the top 25% exceed 356.536 billion IDR.

3.1.2. Distribution Analysis

Histogram and kernel density estimation plots depicting the distribution of AQI, Electricity, and GRDP across Indonesian provinces are illustrated in Figure 2. These visualizations allow for a clear understanding of the spread and skewness of each variable within the dataset. The AQI histogram displays a unimodal distribution with the majority of data points concentrated around the mean, which aligns with the previously noted narrow standard deviation. The kernel density curve suggests that the AQI values are somewhat normally distributed, albeit with a slight leftward skew, indicating a modest number of provinces with lower AQI values. The electricity histogram is markedly right-skewed. This skewness is due to a few provinces with significantly high electricity distribution figures, which pulls the mean to the right of the median. The kernel density plot further accentuates the right tail, emphasizing the inequality in electricity distribution among the provinces. The histogram for GRDP on the right also shows a pronounced right-skewed distribution. A substantial number of provinces have a GRDP that is lower than the mean, while a small number have exceptionally high GRDP values, contributing to the long tail on the right. This pattern is indicative of a few economically dominant provinces, which greatly surpass the others in terms of production or services.

3.1.3. Scatter Plot Analysis

A series of scatter plots comparing the relationships between the following pairs of variables: AQI and Electricity, AQI and GRDP, and Electricity and GRDP across Indonesian provinces are visualized in Figure 3. The first scatter plot illustrates the relationship between AQI and Electricity. The distribution of points is primarily concentrated in the bottom right, suggesting that most provinces experience higher AQI values and lower electricity distribution figures. A few outliers with higher electricity values are observed. The middle scatter plot depicts the relationship between AQI and GRDP. It is evident from the distribution, which is also concentrated in the bottom right, that most provinces have higher AQI values and lower GRDP. Interestingly, the data points indicate that the province with the worst AQI has the highest GRDP, suggesting an inverse relationship between air quality and economic output in this case. The third scatter plot explores the relationship between Electricity and GRDP. This plot shows a clearer clustering towards the bottom left, indicating that many provinces have lower figures for both electricity distribution and GRDP. Additionally, there is a noticeable trend where GRDP increases with increasing electricity distribution among some provinces, as indicated by the cluster of data points stretching towards the top right. This pattern suggests a positive association between these two variables, where provinces with higher electricity distribution tend to have higher economic output.

3.1.4. Box Plot Analysis

The box plots for the AQI, Electricity, and GRDP variables are shown in Figure 4, providing insights into their distribution and identifying any outliers within the Indonesian provinces. The AQI box plot on the left indicates a relatively narrow interquartile range (IQR), suggesting that most of the AQI values are concentrated within a small range, with a few notable outliers indicating exceptionally high AQI values. The middle box plot for electricity shows a wider IQR, reflecting greater variability in electricity distribution across provinces. There are several extreme outliers above the upper whisker, highlighting provinces with particularly high electricity distribution figures. Lastly, the GRDP box plot on the right reveals a similar distribution pattern to electricity, with a wide IQR and several outliers, which points to a substantial spread in the economic output of different provinces and the presence of a few provinces with significantly higher GRDP than the rest.

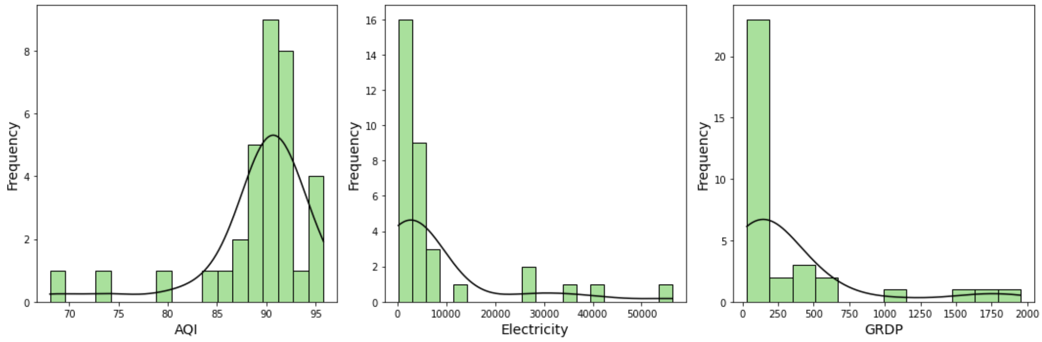

Figure 2. Histogram and kernel density estimation plots for AQI, Electricity, and GRDP across Indonesian provinces.

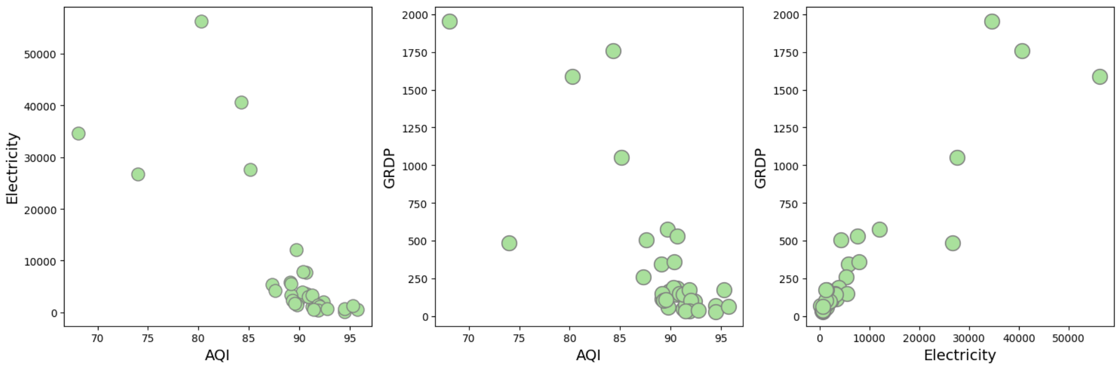

Figure 3. Scatter plots comparing the relationships between the pairs of variables.

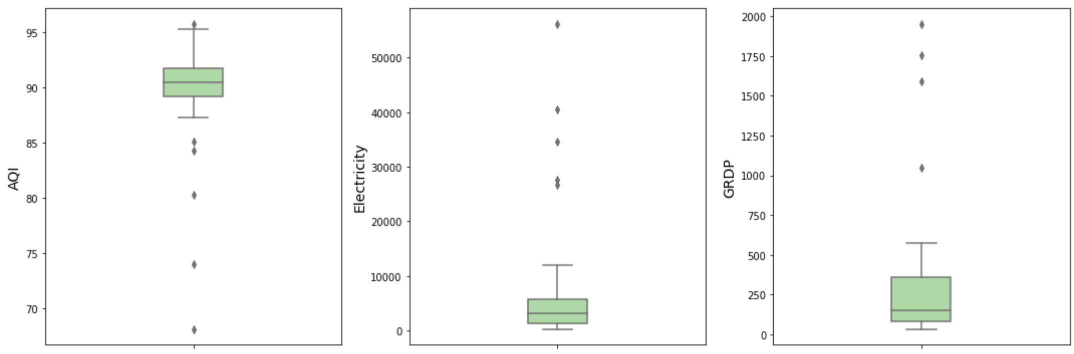

Figure 4. Box plots for the AQI, Electricity, and GRDP variables.

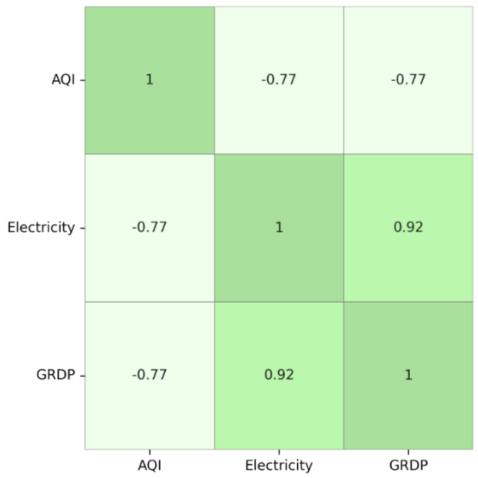

Figure 5. Correlation matrix of the AQI, Electricity, and GRDP variables.

3.1.5. Correlation Analysis

A correlation matrix, providing a visual representation of the relationship between AQI, Electricity, and GRDP within Indonesian provinces, is presented in Figure 5. The values range from -1 to 1, where 1 indicates a perfect positive correlation, -1 is a perfect negative correlation, and 0 is no correlation. Here, AQI has a strong negative correlation with both Electricity and GRDP, as indicated by the value of -0.77. This suggests that higher AQI values are associated with lower electricity distribution and GRDP. Conversely, Electricity and GRDP are highly positively correlated with each other, with a correlation coefficient of 0.92, implying that provinces with higher electricity distribution tend to have higher GRDP values. These strong correlations underscore significant relationships between environmental quality, energy

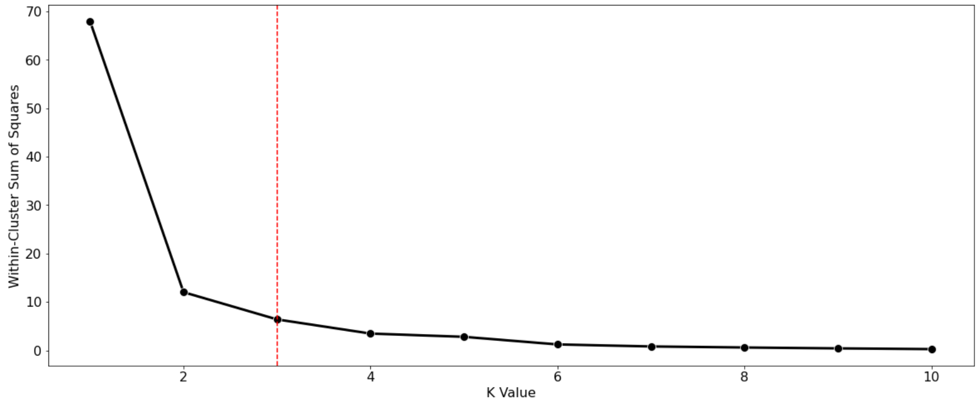

Figure 6. The results of elbow method.

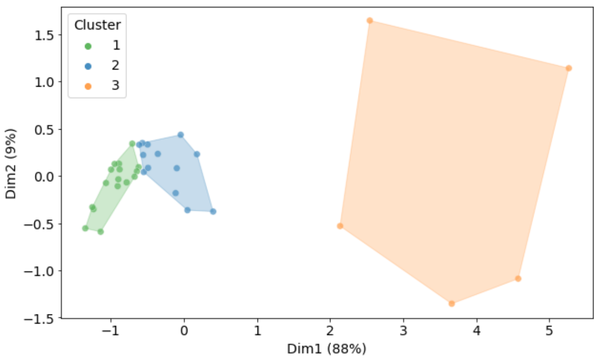

Figure 7. K-Means clustering visualization of Indonesian Provinces.

Table 3. Mean variable values across clusters.

Cluster	AQI	Electricity	GRDP
1	92.363	1399.006	88.992
2	89.437	5058.138	275.426
3	78.352	37124.006	1367.152

distribution, and economic performance across the provinces.

3.2. Clustering Analysis Using K-Means

We utilized the elbow method to determine the optimal value of k. Our analysis covered the range from k=1 to k=10, and the results are depicted in Figure 6. It's evident that a distinct elbow point appears at k=3, indicating that this number of clusters is most suitable for our dataset. The graph illustrates a steep decrease in WCSS from k=1 to k=3, after which the decline stabilizes, suggesting that additional clusters beyond three don't notably enhance cluster homogeneity. Therefore, following the elbow criterion, we chose k=3 as the optimal number of clusters for classifying the Indonesian provinces based on their environmental and economic attributes.

After determining the optimal number of clusters using the elbow method, Figure 7 illustrates the spatial distribution of these clusters among the Indonesian provinces, reduced to two dimensions through Principal Component Analysis (PCA). Each province is represented by a point on the plot and is color-coded according to its assigned cluster from the K-Means clustering algorithm: Cluster 1 in green, Cluster 2 in blue, and Cluster 3 in orange. The two dimensions displayed account for a significant amount of the variance, with dimension 1 explaining 88% and dimension 2 capturing an additional 9%. This reduction into a 2D space while retaining 97% of the variance allows for a clear visual interpretation of the data. The clustering suggests that provinces within the same cluster exhibit similar environmental and economic characteristics.

Upon closer examination of the clusters based on the environmental and economic indicators presented in Table 3, we can discern the distinct characteristics that define each group. Cluster 1 exhibits a moderate mean AQI of 92.363, coupled with a relatively low electricity consumption of 1399.006 gigawatt-hours and a GRDP of 88.992 billion IDR. These figures suggest that provinces in Cluster 1 may have moderate environmental quality and lower economic activity compared to the others. Cluster 2, marked in blue, has a slightly better mean AQI of 89.437, suggesting marginally cleaner air. However, its electricity consumption is significantly higher at 5058.138 gigawatt-hours, and it possesses a larger GRDP of 275.426 billion IDR. This indicates that the provinces in Cluster 2 might have more industrial or economic activity, thus requiring more energy consumption but still maintaining a relatively reasonable environmental standard. Cluster 3, which is indicated by the orange color, stands out with the lowest mean AQI of 78.352, indicating the best air quality among the three groups. It also has the highest mean electricity consumption of 37124.006 gigawatt-hours and a substantially larger GRDP of 1367.152 billion IDR. This could reflect a group of provinces with high levels of economic development and energy consumption, possibly including major urban or industrial centers with better environmental management practices.

Table 4. Clustering of provinces in Indonesia.

Cluster	Provinces
1	Papua Barat, Sulawesi Tengah, Maluku, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Maluku Utara, Aceh, Sulawesi Utara, Kalimantan Utara, Kalimantan Selatan, Kalimantan Barat, Nusa Tenggara Timur, Kep. Bangka Belitung, Bengkulu, Papua
2	Nusa Tenggara Barat, Kep. Riau, Kalimantan Timur, Kalimantan Tengah, Sumatera Utara, Bali, Di Yogyakarta, Sulawesi Selatan, Sumatera Selatan, Sumatera Barat, Riau, Jambi, Lampung
3	Banten, Jawa Tengah, Jawa Timur, Jawa Barat, DKI Jakarta

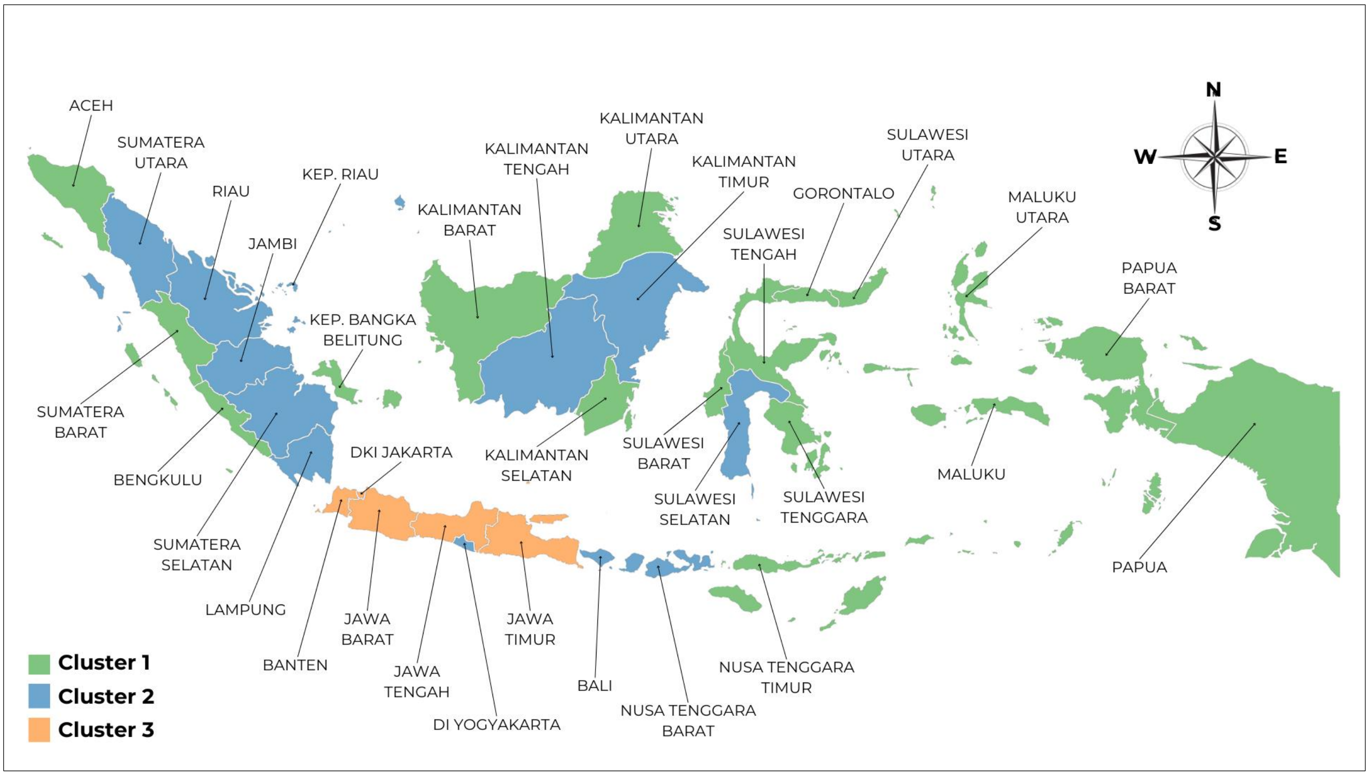

Figure 8. Thematic map of Indonesia shows the geographic distribution of provinces according to clustering results.

Table 4 presents the clustering results of Indonesian provinces obtained through K-Means clustering applied to variables representing environmental and economic factors. Additionally, Figure 8 illustrates a thematic map of the Indonesian archipelago, showing the geographic distribution of clusters derived from the K-Means analysis of environmental and economic data. Cluster 1 consists of a diverse set of provinces, including Papua Barat, Sulawesi Tengah, Maluku, and several others. These provinces are characterized by their vast natural landscapes, which house diverse ecosystems and endemic species. However, they face significant environmental challenges such as deforestation and biodiversity loss, largely driven by unsustainable natural resource exploitation. Economically, these areas rely on primary sectors such as agriculture and mining, which contribute to their lower GDP relative to other regions. The clustering may thus reflect these provinces' shared environmental vulnerabilities and economic profiles, highlighting the need for development strategies that promote environmental sustainability alongside economic growth.

Cluster 2, which includes provinces such as Nusa Tenggara Barat, Kep. Riau, and Lampung, shows a higher level of economic development. Bali's economy, driven by tourism, and Kalimantan Timur's resource-based industries, suggest a mix of economic activities that yield higher overall economic output. However, this economic prosperity comes with substantial environmental costs. It's noteworthy that within the region, the new Indonesian capital city, named Nusantara, is currently under construction. This significant development project symbolizes a strategic economic and administrative shift and is expected to influence the economic landscape of the cluster further. The establishment of Nusantara may lead to infrastructural improvements, increased investment, and economic growth, potentially affecting environmental and economic variables in the future and altering the clustering dynamics.

The third cluster is significantly smaller in terms of the number of provinces, consisting only of Banten, Jawa Tengah, Jawa Timur, Jawa Barat, and DKI Jakarta. These areas are marked by high economic output, driven by manufacturing and services, yet they grapple with the environmental consequences of dense urbanization, such as air and water pollution. The high population density exacerbates these environmental issues, challenging the sustainability of urban centers. The economic and environmental data clustering underscores the need for policies that manage urban growth and improve environmental management in these economically critical regions.

Building on the distinct characteristics identified in each cluster through the K-Means analysis, it becomes imperative to develop customized policy interventions that effectively address the unique economic and environmental challenges faced by these regions. For Cluster 1, policies need to focus on promoting sustainable development and environmental conservation. This might include initiatives such as enhancing sustainable agricultural practices, supporting sustainable forestry management to prevent deforestation, and encouraging investments in renewable energy to diversify the energy mix and reduce ecological footprint.

In Cluster 2, where economic advancement is paired with significant environmental costs, a careful balance between growth and sustainability must be struck, particularly in the context of major developments such as the construction of Nusantara. This situation calls for the implementation of stringent environmental regulations, the adoption of sustainable urban planning practices, and the promotion of green technologies that can mitigate the environmental impacts of rapid development and high economic activity.

Cluster 3, characterized by its densely populated urban centers, demands an integrated approach to urban planning that emphasizes environmental resilience. This could involve revising urban infrastructure to increase green spaces, implementing more rigorous pollution control measures, and investing in smart technologies that enhance efficiency in resource use. These measures are crucial for mitigating the adverse effects of urbanization and supporting sustainable economic growth.

Furthermore, the importance of ongoing monitoring and evaluation cannot be understated. By continually assessing the economic and environmental parameters across these clusters, policymakers can gain valuable insights into the effectiveness of implemented strategies and the evolving dynamics within each cluster. This will enable timely adjustments to policies and interventions, ensuring they remain aligned with both economic objectives and environmental imperatives. This proactive approach is essential for fostering long-term sustainability and resilience in the face of changing economic landscapes and environmental conditions.

4. Conclusions

This study has employed K-means clustering to analyze environmental and economic data from Indonesian provinces in 2022, successfully identifying three distinct clusters that reflect the diverse socio-economic and environmental landscapes across the archipelago. These findings underscore the urgent need for region-specific policies that can drive sustainable development and balance economic growth with environmental conservation.

For provinces with moderate environmental quality and lower economic activity, it is important to enhance agricultural practices, develop eco-tourism, and support renewable energy projects. Provinces with higher economic output and moderate environmental standards should implement smart urban planning, strengthen environmental regulations, and invest in clean technology. In densely populated urban centers with high economic development, expanding green infrastructure, promoting sustainable urban development, and enhancing public transport systems are critical. These targeted interventions will enable Indonesia to address the specific needs of each cluster, promoting a more balanced and sustainable development across its provinces. Continual monitoring and adaptation of these policies will be crucial in responding to dynamic economic conditions and environmental challenges, ensuring long-term sustainability and resilience. This strategic approach will allow Indonesia to harness its full potential while safeguarding its natural resources and enhancing the quality of life for all its citizens.

Author Contributions: Conceptualization, T.R.N. I.H. and G.M.I.; methodology, T.R.N. and N.R.S.; software, N.R.S. and I.S.H.; validation, I.H., Z.Z., R.S. and G.M.I.; formal analysis, T.R.N. and I.H.; investigation, T.R.N., I.H., and N.R.S.; resources, I.H., Z.Z. and G.M.I.; data curation, I.H., N.R.S. and G.M.I.; writing—original draft preparation, T.R.N., Z.Z., R.S. and I.S.H.; writing—review and editing, I.H. and G.M.I.; visualization, T.R.N. and G.M.I.; supervision, G.M.I.; project administration, I.H.; funding acquisition, G.M.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data utilized in this study is available upon reasonable request to the corresponding author.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Herrador, M., and Van, M. L. (2024). Circular Economy Strategies in the ASEAN Region: A Comparative Study, *Science of The Total Environment*, Vol. 908, 168280. doi:10.1016/j.scitotenv.2023.168280.
2. Djais, G., Fransen, J., and Koppenjan, J. (2024). Governing Sustainable Corridor Development: A Case Study of the Gilimanuk–Denpasar–Padang Bai Corridor in Indonesia, *Environmental Policy and Governance*. doi:10.1002/eet.2104.
3. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Enhancing Loan Approval Decision-Making: An Interpretable Machine Learning Approach Using LightGBM for Digital Economy Development, *Malaysian Journal of Computing (MJOC)*, Vol. 9, No. 1, 1734–1745. doi:10.24191/mjoc.v9i1.25691.
4. Kurniadi, A., Weller, E., Salmond, J., and Aldrian, E. (2024). Future Projections of Extreme Rainfall Events in Indonesia, *International Journal of Climatology*, Vol. 44, No. 1, 160–182. doi:10.1002/joc.8321.
5. Tumonggor, M. K., Karafet, T. M., Hallmark, B., Lansing, J. S., Sudoyo, H., Hammer, M. F., and Cox, M. P. (2013). The Indonesian Archipelago: An Ancient Genetic Highway Linking Asia and the Pacific, *Journal of Human Genetics*, Vol. 58, No. 3, 165–173.
6. Wang, X., and Xu, X. (2024). Sustainable Resource Management and Green Economic Growth: A Global Prospective, *Resources Policy*, Vol. 89, 104634. doi:10.1016/j.resourpol.2024.104634.
7. Idroes, G. M., Hardi, I., Hilal, I. S., Utami, R. T., Noviandy, T. R., and Idroes, R. (2024). Economic Growth and Environmental Impact: Assessing the Role of Geothermal Energy in Developing and Developed Countries, *Innovation and Green Development*, Vol. 3, No. 3, 100144. doi:10.1016/j.igd.2024.100144.
8. Idroes, G. M., Hardi, I., Rahman, M. H., Afjal, M., Noviandy, T. R., and Idroes, R. (2024). The Dynamic Impact of Non-renewable and Renewable Energy on Carbon Dioxide Emissions and Ecological Footprint in Indonesia, *Carbon Research*, Vol. 3, No. 1, 35. doi:10.1007/s44246-024-00117-0.
9. Li, Y., Chiu, Y., and Lu, L. C. (2018). Energy and AQI Performance of 31 cities in China, *Energy Policy*, Vol. 122, 194–202. doi:10.1016/j.enpol.2018.07.037.
10. Mahmudah, U., and Lola, M. S. (2023). A Two-Step Cluster for Classifying Provinces in Indonesia Based on Environmental Quality, *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, Vol. 17, No. 3, 1685–1694. doi:10.30598/barekengvol17iss3pp1685-1694.
11. Giao, N. T. (2021). Assessment of Air Quality in Can Tho City, Vietnam Using Cluster Analysis, *Indonesian Journal of Environmental Management and Sustainability*, Vol. 5, No. 4, 154–161. doi:10.26554/ijems.2021.5.4.154-161.
12. Idroes, G. M., Noviandy, T. R., Maulana, A., Zahriah, Z., Suhendrayatna, S., Suhartono, E., Khairan, K., Kusumo, F., Helwani, Z., and Abd Rahman, S. (2023). Urban Air Quality Classification Using Machine Learning Approach to Enhance Environmental Monitoring, *Leuser Journal of Environmental Studies*, Vol. 1, No. 2, 62–68. doi:10.60084/ljes.v1i2.99.
13. Liang, D., Lu, H., Guan, Y., Feng, L., Chen, Y., and He, L. (2023). Further Mitigating Carbon Footprint Pressure in Urban Agglomeration by Enhancing the Spatial Clustering, *Journal of Environmental Management*, Vol. 326, 116715. doi:10.1016/j.jenvman.2022.116715.
14. Zulkepli, N. F. S., Noorani, M. S. M., Razak, F. A., Ismail, M., and Alias, M. A. (2022). Hybridization of Hierarchical Clustering with Persistent Homology in Assessing Haze Episodes between Air Quality Monitoring Stations, *Journal of Environmental Management*, Vol. 306, 114434. doi:10.1016/j.jenvman.2022.114434.
15. Tumiran, T., Budiarto, R., Sarjiya, S., Putranto, L. M., Noorzakiah Naimah, D. Y., Dharmasakya, A. H., and Priyanto, A. (2021). Regional Clustering for Developing Electricity Systems in Archipelagic Area: A Case Study of Maluku and Papua Islands, *2021 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, IEEE, 242–247. doi:10.1109/ICT-PEP53949.2021.9601113.
16. McNeil, M. A., Karali, N., and Letschert, V. (2019). Forecasting Indonesia's Electricity Load through 2030 and Peak Demand Reductions from Appliance and Lighting Efficiency, *Energy for Sustainable Development*, Vol. 49, 65–77. doi:10.1016/j.esd.2019.01.001.
17. Maki, S., Ashina, S., Fujii, M., Fujita, T., Yabe, N., Uchida, K., Ginting, G., Boer, R., and Chandran, R. (2018). Employing Electricity-Consumption Monitoring Systems and Integrative Time-Series Analysis Models: A Case Study in Bogor, Indonesia, *Frontiers in Energy*, Vol. 12, No. 3, 426–439. doi:10.1007/s11708-018-0560-4.
18. Tambunan, H. B., Barus, D. H., Hartono, J., Alam, A. S., Nugraha, D. A., and Usman, H. H. H. (2020). Electrical Peak Load Clustering Analysis Using K-Means Algorithm and Silhouette Coefficient, *2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, IEEE, 258–262. doi:10.1109/ICT-PEP50916.2020.9249773.
19. Arshad, Z., Robaina, M., and Botelho, A. (2020). The Role of ICT in Energy Consumption and Environment: An Empirical Investigation of Asian Economies with Cluster Analysis, *Environmental Science and Pollution Research*, Vol. 27, No. 26, 32913–32932. doi:10.1007/s11356-020-09229-7.
20. Wororomi, J. K., Girik Allo, C. B., and Paranoan, N. R. (2023). Performance of K-Means and DBSCAN Algorithm in Clustering Gross Regional Domestic Product, *Journal of International Conference Proceedings*, Vol. 6, No. 5, 179–193. doi:10.32535/jicp.v6i5.2710.
21. Hardi, I., Ray, S., Attari, M. U. Q., Ali, N., and Idroes, G. M. (2024). Innovation and Economic Growth in the Top Five Southeast Asian Economies: A Decomposition Analysis, *Ekonomikalia Journal of Economics*, Vol. 2, No. 1, 1–14. doi:10.60084/eje.v2i1.145.
22. Idroes, G. M., Syahnur, S., Majid, S. A., Sasmita, N. R., and Idroes, R. (2021). Provincial economic level analysis in Indonesia based on the geothermal energy potential and growth regional domestic products using cluster analysis, *IOP Conference Series: Materials Science and Engineering*, Vol. 1087, No. 1, 012079. doi:10.1088/1757-899X/1087/1/012079.
23. Kurniawan, H., de Groot, H. L. F., and Mulder, P. (2019). Are Poor Provinces Catching-Up the Rich Provinces in Indonesia?, *Regional Science Policy & Practice*, Vol. 11, No. 1, 89–108. doi:10.1111/rsp3.12160.
24. Sasmita, N. R., Phonna, R. A., Fikri, M. K., Khairul, M., Apriliansyah, F., Idroes, G. M., Puspitasari, A., and Saputra, F. E. (2023). Statistical Assessment of Human Development Index Variations and Their Correlates: A Case Study of Aceh Province, Indonesia, *Grimsa Journal of Business and Economics Studies*, Vol. 1, No. 1, 12–24.
25. Maulidar, P., Fitriyani, F., Sasmita, N. R., Hardi, I., and Idroes, G. M. (2024). Exploring Indonesia's CO2 Emissions: The Impact of Agriculture, Economic Growth, Capital and Labor, *Grimsa Journal of Business and Economics Studies*, Vol. 1, No. 1, 43–55. doi:10.61975/gjbes.v1i1.22.
26. Ahmed, M., Seraj, R., and Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance

Evaluation, *Electronics*, Vol. 9, No. 8, 1295. doi:10.3390/electronics9081295.

27. Sinaga, K. P., and Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm, *IEEE Access*, Vol. 8, 80716–80727. doi:10.1109/ACCESS.2020.2988796.

28. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., and Heming, J. (2023). K-Means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data, *Information Sciences*, Vol. 622, 178–210. doi:10.1016/j.ins.2022.11.139.

29. Surya, B., Salim, A., Hernita, H., Suriani, S., Menne, F., and Rasyidi, E. S. (2021). Land Use Change, Urban Agglomeration, and Urban Sprawl: A Sustainable Development Perspective of Makassar City, Indonesia, *Land*, Vol. 10, No. 6, 556. doi:10.3390/land10060556.

30. Noviandy, T. R., Maulana, A., Sasmita, N. R., Suhendra, R., Muslem, M., Idroes, G. M., Paristiowati, M., Helwani, Z., Yandri, E., Rahimah, S., Muhammad, M., Irvanizam, I., and Idroes, R. (2020). The Implementation of K-Means Clustering in Kovats Retention Index on Gas Chromatography, *IOP Conf. Ser.: Mater. Sci. Eng*.

31. Kramer, O. (2016). Scikit-Learn, 45–53. doi:10.1007/978-3-319-33383-0_5.

32. Arthur, D., and Vassilvitskii, S. (2007). K-means++: The Advantages of Careful Seeding, *Soda* (Vol. 7), 1027–1035.

33. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.