ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

"From Data to Impact: Leveraging Innovation in Data Science and Official Statistics for Data-Driven Policy Making and Global Competitiveness"

Clustering of Junior High School Education in West Java Based on Density and Dropout Ratios Using Quartile and K-Means Methods

E Nurkhofifah[1] *, D Athina[2], A R Tarida[3], F A Pratiwi[3]

[1] Computer Science, Universitas Pendidikan Indonesia, Bandung, Indonesia
[2] PT Astra International Tbk, Jakarta, Indonesia
[3] Pusat Data dan Teknologi Informasi, Kementerian Pendidikan Dasar dan Menengah, Tangerang Indonesia

*Corresponding author's email: evaanurkhofifah@gmail.com

Abstract. Education disparities across regions often reflect differences in school density, teacher availability, and student dropout rates. This study aims to classifies junior high school education in West Java into more homogeneous groups to better understand these disparities. Two clustering approaches were applied: quartile grouping and the K-Means algorithm. Quartile grouping provided a simple categorization of each indicator into four levels (very high, high, low, very low), while K-Means offers a more flexible and data-driven segmentation. K-Means algorithm produced three distinct clusters: (1) Balanced and Stable regions with proportional ratios and low dropout rates, (2) High-Density but Stable regions concentrated in urban and peri-urban areas with high student-teacher and student-school ratios but controlled dropout levels, and (3) Elevated Dropout Risk regions, mostly in rural and southern areas, with lower density but higher dropout rates. The comparison shows that quartile grouping is easy to interpret for individual indicators, while K-Means provides more comprehensive insights into multidimensional patterns. This research highlights the potential of clustering methods to guide policymakers in designing differentiated strategies, from infrastructure expansion in dense regions to social support programs in dropout-prone areas.

Keyword: Clustering Analysis, Education Inequality, K-Means, Quartile Grouping, West Java

1. Introduction

Educational equity remains a prominent challenge in Indonesia's education system, particularly at the lower secondary level, as persistent disparities in infrastructure, access, and teacher allocation continue to undermine its effectiveness [1][2]. This level represents a crucial transition point from primary to secondary education, meaning that disparities at this stage may significantly affect students' continuation to higher levels of education [3]. As the most populous province in Indonesia, West Java faces a substantial educational burden, with highly diverse distributions of schools, teachers, and students across its regions.

According to BPS (2024), the School Participation Rate (APS) for children aged 13-15 years (lower secondary level) in West Java reached 95.86%, indicating that around 4-5% of children at this age are still out of school. Although this figure reflects relatively high access to lower secondary education, it also highlights the existence of marginalized groups. The challenge becomes more evident when considering the APS for children aged 16-18 years (upper secondary level), which is only 71.25%, meaning that nearly 29% of adolescents at the upper secondary age do not continue their formal education. This gap underscores that while access at the Junior High School level is relatively equitable, the transition to upper secondary remains a significant issue.

Regional disparities in educational development persist, as evidenced by inadequate school facilities and an insufficient ratio of teachers to schools in several areas across Indonesia [4][5]. Inequalities in teacher distribution, facilities, and student allocation across densely populated areas in West Java also persist as critical concerns. Disparities in student numbers across schools, non-ideal student–teacher ratios, and dropout rates serve as key indicators to assess the effectiveness and efficiency of educational delivery. To better understand these conditions, it is necessary to analyze not only each indicator separately but also the interactions among them to identify broader regional patterns. As argued in [6], the choice of indicators and analytical approach inherently reflects normative assumptions about educational equity and justice.

Conventional approaches, such as quartile grouping, are often applied to identify priority areas in education planning due to their simplicity. Previous research employed the quartile grouping method to classify private higher education institutions into four performance categories: poor, fair, good, and excellent, based on their KPI achievements [7]. However, this method has limitations in capturing the natural clustering of multidimensional indicators. An alternative and more exploratory approach is clustering analysis, such as the K-Means algorithm, which groups regions based on similarities across multiple characteristics in an objective manner. As surveyed in [8], clustering techniques in educational must adhere not only to quality metrics but also fairness constraints, particularly when used for sensitive applications such as student grouping. The review highlights evaluation benchmarks, fairness-aware models, and challenges such as scalability and high-dimensional datasets.

Several previous studies have employed clustering techniques in educational contexts to uncover patterns of disparity or performance. Previous studies have applied K-Means clustering to educational regions, but only based on student-teacher ratios and limited to a single district[9]. One key study conducted in West Java applied and compared K-Means, K-Medoids, and Fuzzy C-Means to classify school accreditation in West Java, finding K-Medoids to be the most effective based on clustering validity indices [10]. In another context, in [11] utilized K-Means clustering across Indonesian provinces, involving eleven parameters such as schools, students, dropouts, teachers, classrooms, and more. Grouping West Java and East Java into similar performance clusters. A previous study applied K-Means and DBSCAN clustering algorithms to categorize Indonesian cities based on dropout rates across education levels [12]. Another study conducted educational mapping in West Java Province using Agglomerative Hierarchical Clustering (AHC) with various linkage methods, finding disparities in the distribution of schools and teachers, particularly at the upper secondary and tertiary levels, based on cophenetic correlation coefficients [13].

This study investigates regional educational disparities in West Java by applying both quartile grouping and K-Means clustering on three core indicators: student-school ratio, student-teacher ratio, and dropout rate. These indicators were selected because they represent critical dimensions of educational access, resource allocation, and student retention, which are central to understanding inequality in the Indonesian education system. While quartile grouping offers a straightforward categorization, K-Means clustering facilitates the identification of latent patterns and complex regional groupings. This study fills this gap by employing both quartile grouping and K-Means clustering to assess educational inequality patterns in West Java. By using both traditional and data-driven classification methods, the study provides complementary perspectives that enhance the robustness of the analysis. The resulting clusters offer actionable insights into regional disparities and serve as an empirical basis for formulating more targeted and equitable education policies. Ultimately, the study seeks to contribute to the advancement of educational equity through data-driven insights grounded in both statistical rigor and ultimately supporting the development of more targeted education policies.

2. Research Method

This study employed a quantitative research design with a secondary data approach. The analysis focused on numerical indicators of educational equity, which were subsequently processed using descriptive statistics and clustering techniques. To ensure clarity of the research process, the methodological framework is illustrated in the figure 1, which outlines the sequence from data collection to analytical procedures.

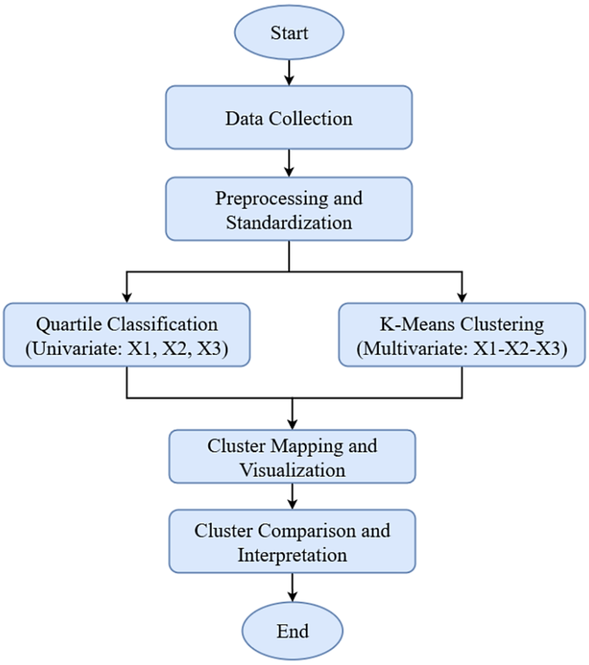

Start
Data Collection
Preprocessing and Standardization
Quartile Classification (Univariate: X1, X2, X3)
K-Means Clustering (Multivariate: X1-X2-X3)
Cluster Mapping and Visualization
Cluster Comparison and Interpretation
End

Figure 1. Methodological Framework of the Study.

This study with a focus on analysis to examine educational disparities at the junior secondary school level in West Java Province, using data from the 2024 education dataset published by the Ministry of Education (Kemendikdasmen). The unit of analysis is the district/city (Kabupaten/Kota), covering all administrative areas in the province. Three key indicators were selected: (X1) Student-School Ratio, defined as the number of students per school in each district; (X2) Student-Teacher Ratio, calculated as the number of students per teacher; and (X3) Dropout Rate, expressed as the percentage of students who did not complete junior secondary education. These variables were chosen to reflect three core dimensions of educational equity: access, resource distribution, and retention. To analyze disparities based on these indicators, the study employed two techniques: a quartile-based approach for initial distributional assessment, and K-Means clustering to identify groups of districts with similar educational characteristics. This structured approach ensures methodological rigor and enables a more comprehensive understanding of educational inequality across districts/cities in West Java.

2.1. Data Collection

The dataset was obtained from the official open data portal of the Indonesian Ministry of Education (Kemendikdasmen), which provides comprehensive administrative education data for the year 2024 at the sub-national level. The data were retrieved in tabular format and subsequently processed and cleaned using Python within the Google Colab environment. Three core variables were extracted for analysis: (1) the student-school ratio, defined as the total number of enrolled students divided by the number of operational schools; (2) the student-teacher ratio, calculated as the number of students per active teacher; and (3) the dropout rate, measured as the percentage of students who discontinued formal schooling within a given academic year. These indicators were selected as proxies for key dimensions of educational equity: access (student-school ratio), resource distribution (student-teacher ratio), and retention (dropout rate).

2.2. Preprocessing and Standardization

Pre-processing involved consistency checks and the calculation of derived ratios to ensure comparability across regions. The dataset covered all 27 districts/cities in West Java for the year 2024 and contained no missing values across the three variables. Accordingly, no case deletion or imputation was required. Initial verification was performed manually using spreadsheet tools (e.g., Google Sheets) to inspect formatting and confirm the accuracy of aggregate values provided. This was followed by programmatic checks in Python to standardize and compute the ratios (student-school, student-teacher, and dropout rate).

After the cleaning process, the dataset was accessed directly in Python using a public CSV export link from Google Sheets. This method allowed for seamless integration with Python-based data processing workflows without requiring file downloads. Within the Python environment, selected numerical variables were normalized using the StandardScaler method from the scikit-learn library. This transformation standardized the features to have a mean of zero and a standard deviation of one, ensuring that variables with different units and scales contributed equally to the clustering process. This hybrid approach allowed for both manual precision during data validation and efficient, replicable preprocessing steps required for unsupervised machine learning techniques such as K-Means clustering.

2.3. Classification and Clustering Techniques

To identify regional patterns of educational inequality, this study employed two complementary analytical techniques: quartile-based classification and K-Means clustering. The quartile method was used as an initial exploratory step to describe the distribution of each educational indicator across districts, allowing for a straightforward comparison based on relative rankings. In contrast, K-Means clustering was applied as an unsupervised learning algorithm to group districts with similar educational profiles based on multiple variables simultaneously. Together, these techniques enabled both descriptive insights and data-driven segmentation, strengthening the interpretation of disparities across regions.

Each variable was first examined using descriptive statistics, including the mean, minimum, maximum, and standard deviation, to provide a general overview of distribution and variation across districts. Following this, a univariate quartile grouping was applied to each variable individually: X1 (student-school ratio), X2 (student-teacher ratio), and X3 (dropout rate). This approach enabled the identification of relative rankings and positional inequalities for each educational indicator. By dividing regions into four groups (Q1-Q4) based on statistical distribution cut-offs, providing a straightforward prioritization scheme. It also allows for the identification of high-need areas for each variable independently. The quartile thresholds were calculated using the following formula:

$$Q_k = x\left(\frac{k(n+1)}{4}\right) \quad (1)$$

where x is the sorted data, n is the total number of observations, and k=1,2,3 corresponds to the lower, median, and upper quartile positions.

In addition to the univariate classification, a multivariate clustering technique, K-Means clustering was employed to group districts based on the combined values of all three variables (X1, X2, and X3). This unsupervised machine learning method partitions observations into a predefined number of clusters by minimizing within-cluster variance. Prior to clustering, all variables were standardized to zero mean and unit variance using the StandardScaler method, ensuring equal contribution of each variable during distance calculation. The standardization follows the formula:

$$Z_{ij} = \frac{X_{ij}-\mu_j}{\sigma_j}, \ j = \in \{X_1, X_2, X_3\} \quad (2)$$

where Z is the standardized value, X is the original value, μ is the mean, and σ is the standard deviation of each variable.

To determine the optimal number of clusters (k), this study employed three internal validation techniques commonly used in unsupervised learning: the Elbow Method, the Silhouette Score, and the Davies-Bouldin Index (DBI). The Elbow Method identifies the point where adding more clusters yields diminishing returns in reducing within-cluster sum of squares (WCSS). The Silhouette Score evaluates the cohesion and separation of clusters, with higher values indicating more well-defined groupings. The DBI considers both intra-cluster similarity and inter-cluster separation, where lower scores reflect more compact and distinct clusters. While these metrics offer a robust initial assessment of clustering quality, the study did not implement further stability or robustness checks (e.g., bootstrapping, repeated sampling). This is noted as a methodological limitation due to data and scope constraints, and future research is encouraged to incorporate additional validation strategies for more comprehensive evaluation.

Following the determination of the optimal number of clusters, the main clustering was conducted using the K-Means algorithm, an unsupervised learning method designed to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (centroid). This technique aims to minimize the total WCSS, commonly referred to as inertia.

$$arg\ min_C \sum_{i=1}^{k} \sum_{x \in C_i} |x - \mu_i|^2 \quad (3)$$

where Ci is the set of points assigned to cluster i, and μ_i is the centroid of cluster i. The algorithm proceeds iteratively through two main steps: (1) assigning each data point to the nearest centroid using Euclidean distance, and (2) updating the centroids by calculating the mean of all points within each cluster. This process continues until the cluster assignments no longer change or a maximum number of iterations is reached.

In this study, clustering was implemented using Python's scikit-learn library with the KMeans class. All variables were previously standardized using StandardScaler to ensure comparability in scale. The number of initializations (n_init) was set to 10 to reduce the risk of convergence to local minima, and a fixed random_state was used to ensure reproducibility. This clustering process enabled the identification of region-level. Each cluster was profiled and labeled according to dominant characteristics, translating statistical results into actionable categories for policy interpretation. Clusters were labeled according to their composite features, such as high dropout rates, student-teacher imbalances, or relatively stable education indicators. The labeling process allowed for a clearer translation of numeric clustering results into actionable policy categories, enabling stakeholders to differentiate between types of need such as capacity issues versus retention challenges across regions.

K-Means was chosen for its computational efficiency, ease of interpretation, and suitability for low-dimensional numerical data such as the one used in this study. While more advanced clustering algorithms (e.g., K-Medoids, DBSCAN, or hierarchical clustering) may provide alternative perspectives, the study's focus was on applying a widely adopted, interpretable method to uncover spatial disparities in education. The comparative use of K-Means and quartile-based grouping also highlights the strengths and limitations of different classification strategies in the context of regional education analysis. The limited scope of clustering methods is acknowledged as a trade-off for clarity and focus in this policy-oriented study and is suggested as a direction for future research.

2.4. Cluster Mapping and Visualization

To visualize the geographic distribution of the clusters across West Java, this study utilized spatial mapping based on a GeoJSON file containing the administrative boundaries of all districts and cities in the province. The cluster labels (0, 1, 2), assigned during the K-Means clustering process, were merged with the geographic dataset using district-level administrative codes as unique identifiers.

Spatial visualization was implemented in Python using the geopandas and folium libraries. Each region was colored according to its cluster label, enabling a thematic map that clearly illustrated spatial patterns in educational disparities. This visualization technique not only aids in interpretation but also provides a policy-relevant tool for identifying regional clusters at risk, such as areas with elevated dropout rates or overcrowded school environments. By integrating statistical clustering with geospatial data, the study ensures that the analytical findings can be operationalized for region-specific interventions. The maps serve as both a diagnostic and planning tool for stakeholders aiming to address inequality across the educational landscape.

3. Result and Discussion

This section presents the findings of the study and their interpretation in relation to the research objectives. The analysis begins with descriptive statistics to provide an overview of the key variables, followed by univariate classification using quartiles. Subsequently, multivariate clustering with the K-Means algorithm is applied to capture more complex patterns across regions. Finally, the results are compared and discussed in the context of educational participation and policy implications.

3.1. Result

The initial descriptive analysis provides an overview of the distribution of the three key variables: student–school ratio, student–teacher ratio, and dropout rate. Descriptive statistics for the three variables are presented in table 1.

Table 1. Descriptive statistics of education ratio variables.

Descriptive Statistics	X_1	X_2	X_3 (%)
Mean	320.12	19.44	0.053
Standard Deviation	60.26	2.50	0.045
Q1	288.05	17.57	0.023
Median	328.31	19.79	0.047
Q3	345.87	20.91	0.063
Min	208.89	14.94	0.000
Max	465.17	25.43	0.205

Based on table 1, the student-school ratio (X1) has a mean of 320.12 students per school, with a minimum value of 208.89 and a maximum of 465.17. This indicates disparities in student distribution and school capacity across regions. The student-teacher ratio (X2) ranges from 14.94 to 25.43, with an average of 19.44. This suggests that, on average, each teacher is responsible for approximately 19 to 20 students, although there is considerable variation among regions. The dropout rate (X3) is relatively low, with a mean value of 0.053%. However, some regions exhibit dropout rates as high as 0.205%, highlighting that although dropout cases are rare, they still warrant policy attention. The dropout was calculated as the proportion of students discontinuing school during the 2024 academic year relative to total enrollment at the beginning of the year. All values are expressed in percentages (×100) for clarity. Furthermore, quartile values for each indicator will be utilized in the subsequent quartile-based regional clustering. The observed variability, particularly in the student-school ratio and the dropout rate, suggests potential differentiation among regions in terms of educational burden and challenges. Univariate analysis with quartiles, each variable was examined separately using quartile grouping. Quartile values (Q1, Q2, Q3) for each indicator were then used in quartile-based regional clustering, classifying districts and municipalities into four categories, very low, low, high, and very high, to provide an initial overview of educational burdens and regional disparities.

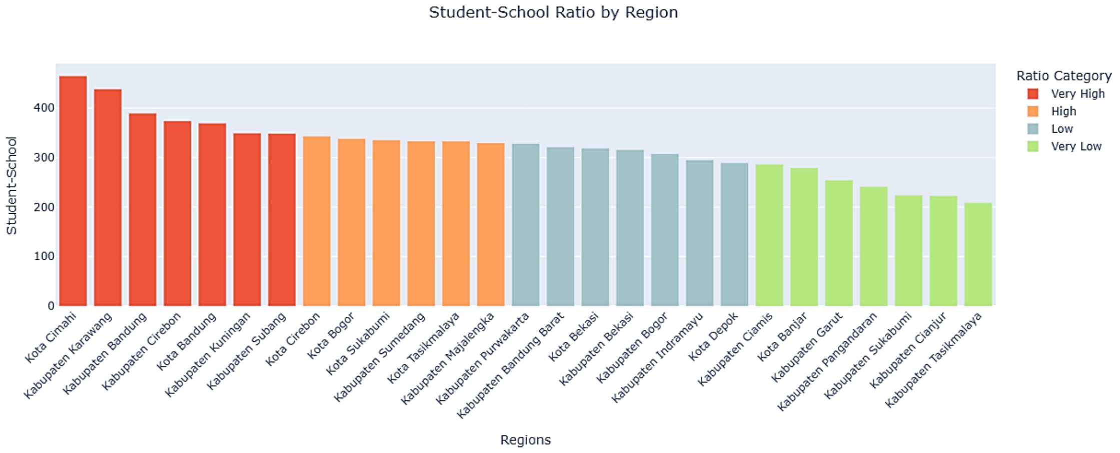


Figure 2. District-level student-school ratios by quartile category.

The quartile grouping in figure 2 shows that several urban and semi-urban regions such as Kota Cimahi, Kabupaten Karawang, Kabupaten Bandung, and Kota Bandung fall into the very high category. This indicates a relatively dense concentration of students in each school, which often reflects either rapid population growth or limited expansion of school infrastructure. On the other hand, regions such as Kabupaten Ciamis, Kota Banjar, and Kabupaten Garut are in the very low category, suggesting that the number of schools in these areas is relatively adequate compared to the student population. These differences highlight the imbalance between school capacity and population density across West Java.

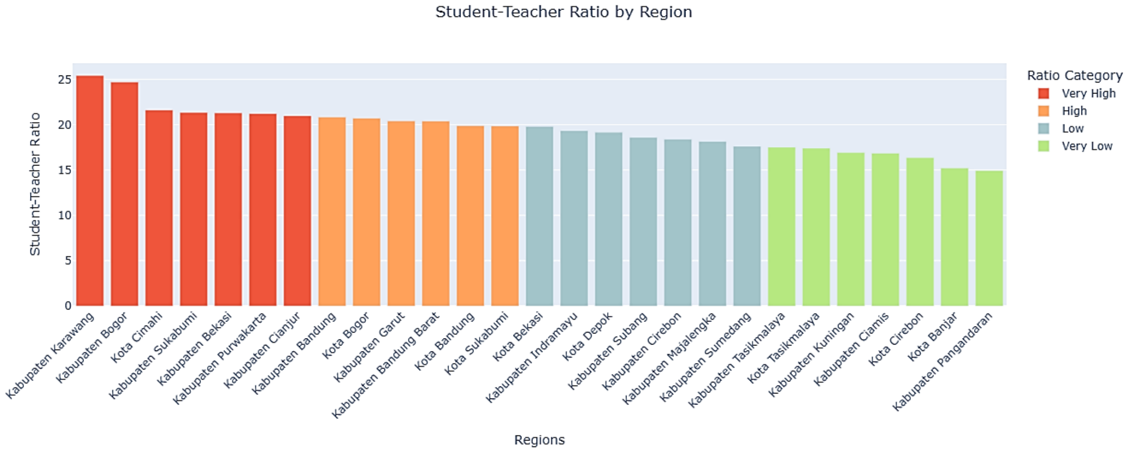


Figure 3. District-level student-teacher ratios by quartile category.

In terms of teacher availability in figure 3, Kabupaten Karawang, Kabupaten Bogor, and Kabupaten Bekasi are classified as very high, meaning each teacher is responsible for a large number of students. This condition often occurs in regions with rapid population growth but slower recruitment or distribution of teachers. Conversely, Kabupaten Tasikmalaya, Kota Tasikmalaya, and Kabupaten Kuningan are placed in the very low quartile, implying a more favorable distribution of teachers relative to the number of students. Such variation reflects disparities in teacher allocation, which may be influenced by geographic accessibility and local education policies.

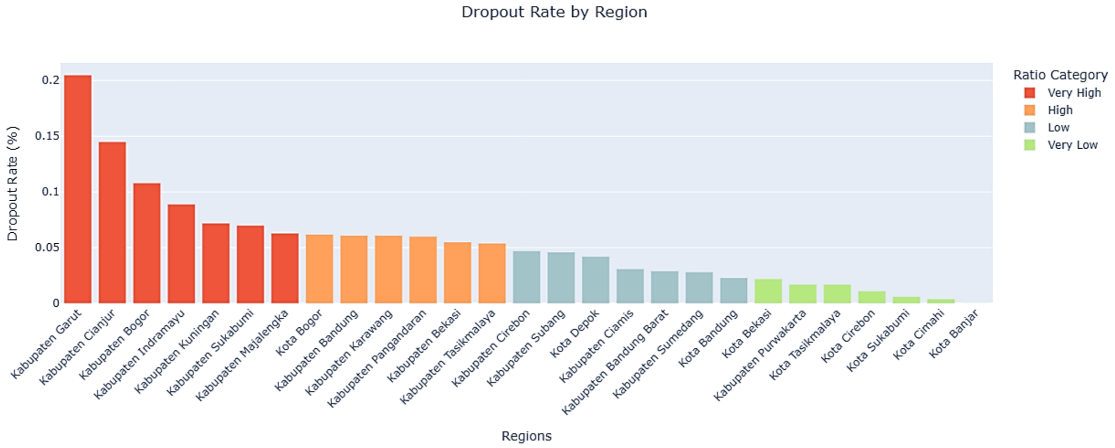


Figure 4. District-level dropout rate by quartile category.

The dropout rate in figure 4, analysis reveals that Kabupaten Garut, Kabupaten Cianjur, and Kabupaten Bogor are in the very high category. These areas are often characterized by socio-economic challenges and limited accessibility, which contribute to higher dropout levels. In contrast, urban areas such as Kota Bekasi, Kota Tasikmalaya, and Kota Cimahi are classified as very low, suggesting better access to educational facilities and stronger household support for continued schooling. The presence of urban infrastructure and higher socio-economic capacity in these cities may reduce dropout risks compared to rural districts.

Urban and peri-urban areas generally occupy the high or very high quartiles for student-school and student-teacher ratios, reflecting population concentration and rapid urban growth that outpace infrastructure and teacher availability. In contrast, rural or peripheral regions tend to fall into lower quartiles, indicating smaller school-age populations but not necessarily stronger educational capacity. These quartile findings reveal disparities in resources and the interaction between structural conditions and educational outcomes, urban centers face high density yet manage to maintain low dropout rates, whereas rural regions with limited resources experience compounded disadvantages, underscoring the need for redistributive policies and strategic teacher deployment. While the quartile method provides a simple distribution-based classification, clustering techniques offer a more data-driven approach to identify groups with similar characteristics. Accordingly, correlation and PCA analyses (Figures 5 and 6) were employed to examine inter-indicator relationships and justify the joint use of the three indicators in the subsequent K-Means clustering.

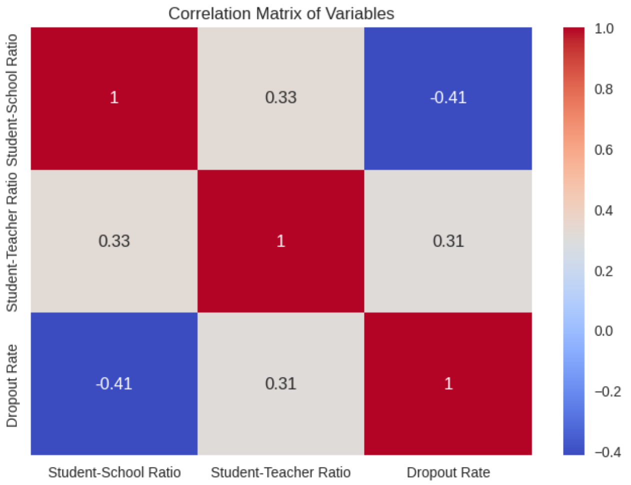


Figure 5. Correlation Matrix of Variables.

Based on the correlation matrix in figure 5, the student-teacher ratio and the student-school ratio exhibit a weak positive correlation (r = 0.33), indicating that regions with larger schools tend to have slightly higher student–teacher ratios. Similarly, the student-teacher ratio and the dropout rate are weakly correlated (r = 0.31). In contrast, the student-school ratio shows a moderate negative correlation with the dropout rate (r = -0.41), suggesting that regions with more crowded schools may not necessarily experience higher dropout, and in some cases even the opposite. Overall, the correlations are relatively low, implying that the three indicators capture distinct dimensions of educational conditions, thereby justifying their joint use in clustering analysis.

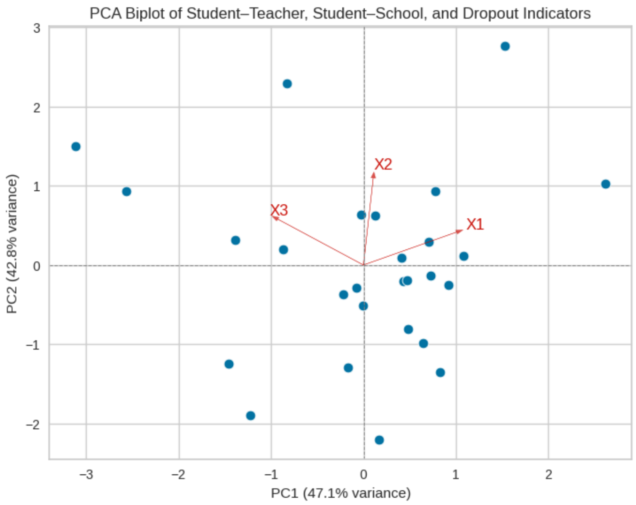


Figure 6. PCA biplot of education ratios.

The PCA biplot in figure 6 shows that the first two principal components explain 89.9% of the total variance (PC1 = 47.1%, PC2 = 42.8%). The student-teacher ratio (X2) contributes strongly to PC1, while the student-school ratio (X1) loads more heavily on PC2. The dropout rate (X3) also aligns primarily with PC2 but in the opposite direction to X1, indicating a contrasting pattern between school crowding and dropout. The clear separation of variable vectors suggests that the three indicators capture distinct dimensions of educational conditions, supporting their joint use in clustering analysis.

The optimal number of clusters was determined using the Elbow method, which identifies the point where adding more clusters no longer significantly reduces within-cluster variance, and the Silhouette Score, which evaluates how well each data point fits its cluster relative to others, with scores closer to 1 indicating better separation.

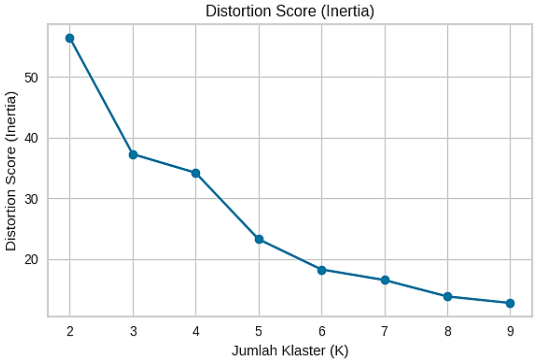


Figure 7. Elbow method graph.

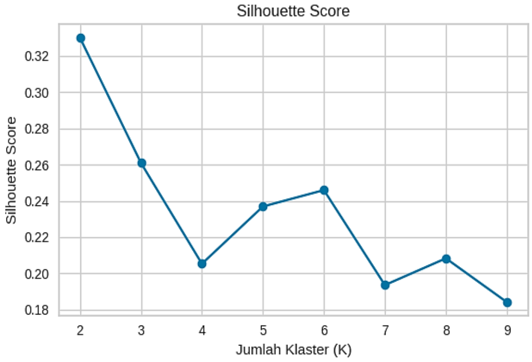


Figure 8. Silhouette score graph.

Figure 7 shows the Elbow curve with a clear decline up to three clusters (k = 3), while Figure 8 indicates the highest Silhouette Score at k = 2 (0.33), though the score for k = 3 (0.261) remains acceptable. The Elbow method and the DBI, which decreases from 1.225 (k = 2) to 1.050 (k = 3), both support selecting three clusters for better compactness and separation. Validation with standardized data and multiple initializations (n_init = 10) confirms improved clustering stability compared to a single initialization.

Table 2. Cluster Validation Metrics for Different k.

k	WCSS	Silhouette	CH	DBI
2	54.761	0.354	11.979	1.226
3	37.192	0.264	14.134	1.049
4	27.928	0.278	14.569	0.988
5	22.133	0.286	14.628	0.939

The extended validation results in table 2 with standardized data and multiple initializations confirm this choice. The within-cluster sum of squares (WCSS) decreases as the number of clusters increases, as expected. The Silhouette coefficient reaches its highest value at k=2 (0.354), but the difference compared to k=3 (0.264) is relatively small. Similarly, the Calinski-Harabasz index remains within a narrow range (around 14) for k=3 to k=5, and the Davies-Bouldin index also shows only minor differences across k=2 to k=5. Given these small variations across the metrics, the decision of the optimal k cannot rely solely on numerical superiority. Choosing k=3 provides a more balanced solution it avoids the oversimplification of only two broad groups while preventing the instability of very small clusters observed at k=4 and k=5 containing only two members. Thus, k=3 is considered the most representative clustering configuration, balancing interpretability and statistical validity.

The next stage applied the K-Means algorithm to group regencies/cities based on similar patterns across the three educational variables: student-teacher ratio, student-school ratio, and dropout rate. Using the predetermined k value, 27 regencies/cities in West Java were segmented into three clusters, each reflecting distinct educational conditions. In table 3 shows that each cluster represents areas with unique educational challenges and characteristics.

Table 3. Cluster summary statistics and regional composition.

Cluster	Indicator	Mean	Median	IQR	Regions
0 (Balanced and Stable)	X1 X2 X3	303.196 17.341 0.043	312.446 17.465 0.044	51.135 1.527 0.036	Kabupaten Sumedang, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Majalengka, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Pangandaran, Kota Cirebon, Kota Depok, Kota Tasikmalaya, Kota Banjar.
1 (High-Density but Stable)	X1 X2 X3	363.194 20.857 0.035	338.484 20.710 0.029	56.861 1.391 0.038	Kabupaten Bandung, Kabupaten Cirebon, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kota Bandung, Kota Bogor, Kota Sukabumi, Kota Bekasi, Kota Cimahi.
2 (Elevated Dropout Risk)	X1 X2 X3	252.429 21.858 0.132	239.538 21.165 0.126	43.893 1.351 0.061	Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut.

Table 3 summarizes the statistical characteristics of each cluster based on central tendency and variability measures. Cluster 0 is characterized by moderate student-school ratios, low student-teacher ratios, and low dropout rates, indicating relatively balanced educational conditions in less dense areas such as Sumedang, Tasikmalaya, and Depok, generally represents regions outside the major urban centers, where population pressure is lower but education systems remain relatively balanced. Cluster 1 shows the highest student-school ratios and moderately high student-teacher ratios but maintains low dropout rates, particularly among urban regions with densely populated and urban areas such as Bandung City and Bekasi City, reflecting efficient education systems supported by better infrastructure and resources in urban settings. In contrast, Cluster 2 exhibits lower student density but higher dropout rates and teacher shortages which includes Garut and Cianjur and surrounding areas, suggesting more severe educational challenges in less developed regions. To further explore the relationships among the indicators and visualize the separation between clusters, Principal Component Analysis (PCA) was conducted. The PCA biplot presented in figure 9 illustrates these patterns and supports the interpretation derived from the summary statistics.

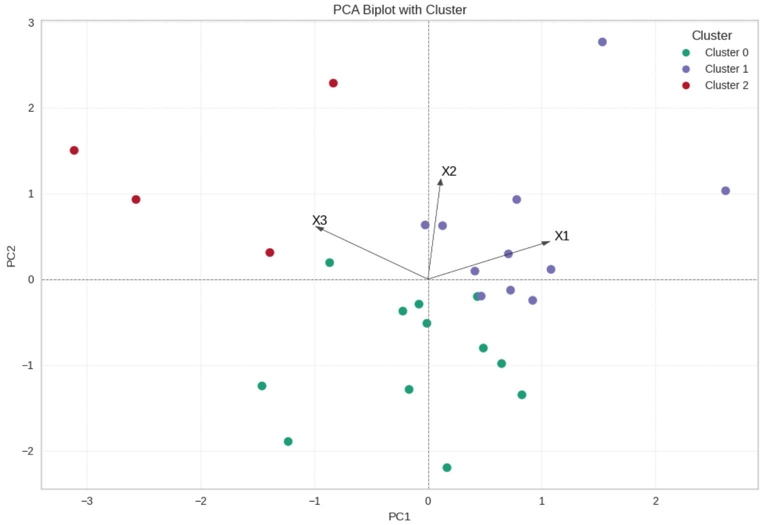
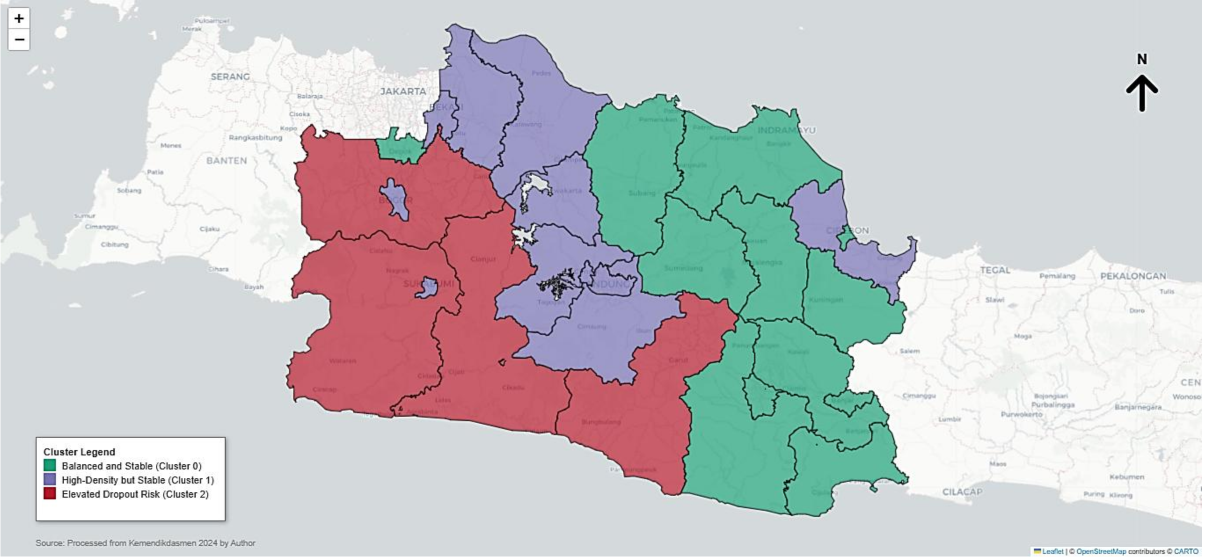


Figure 9. PCA with cluster.

The PCA biplot in figure 9 shows clear separation among the three clusters based on key educational indicators. Cluster 0 (green) appears in the lower-left quadrant with moderate student-school (X1) and low student-teacher ratios (X2), Cluster 1 (purple) in the upper-right with higher X1 and moderate X2, and Cluster 2 (red) on the far left, characterized by high dropout rates (X3) and low X1-X2 values. The arrow directions indicate that X1 and X2 contribute positively to PC1 and PC2, while X3 contributes negatively along PC1, distinguishing high-risk from stable regions. Figure 10 further maps these clusters spatially, showing distinct regional patterns consistent with the PCA and statistical findings.


+
−
N
SERANG
JAKARTA
BANTEN
BEKASI
BOGOR
SUKABUMI
BANDUNG
INDRAMAYU
CIREBON
TEGAL
PEKALONGAN
CILACAP
Cluster Legend
🟩 Balanced and Stable (Cluster 0)
🟪 High-Density but Stable (Cluster 1)
🟥 Elevated Dropout Risk (Cluster 2)
Leaflet | © OpenStreetMap contributors © CARTO

Source: Processed from Kemendikdasmen 2024 by Author

Figure 10. Results of clustering West Java regions based on education ratio using KMeans.

Figure 10 illustrates the spatial distribution of clusters generated by the K-Means algorithm. Green indicates areas belonging to Cluster 0 (Balanced and Stable), purple represents Cluster 1 (High-Density but Stable), and red shows Cluster 2 (Elevated Dropout Risk). The visualization reveals that green cluster areas (Balanced and Stable) are fairly evenly distributed across the central and eastern parts of West Java, covering many regions outside major urban centers. The purple cluster (High-Density but Stable) tends to be concentrated in urban and peri-urban areas, particularly around the Greater Bandung area, Bogor, Depok, Bekasi, and the western part of the northern coastal region. Meanwhile, the red cluster appears concentrated in the southern and southwestern parts of the province, areas generally characterized by more challenging geographic conditions and more complex socio-economic factors. These characteristics reinforce the label "Elevated Dropout Risk", and imply that interventions in these areas should focus not only on education personnel redistribution but also on broader social support programs to improve student retention.

Overall, this spatial mapping confirms that the cluster distribution is not random, but reflects structural regional disparities. This insight reinforces the need for spatially targeted policy interventions, tailored to the distinct challenges of each cluster. While the spatial distribution offers visual insights into regional disparities, statistical testing in table 4 was carried out to examine whether these differences are also significant in terms of educational indicators. Although the variables tested are the same ones used in forming the clusters, the Kruskal-Wallis test serves as a complementary statistical validation. Rather than seeking new findings, this test confirms that the clusters formed through K-Means represent statistically meaningful groupings, thus supporting the internal consistency and interpretability of the clustering results.

Table 4. Kruskal–Wallis and Post-hoc Dunn Test Results.

Indicator	H (Kruskal–Wallis)	p-value	ε^2	Significant Pairwise Differences (Post-hoc Dunn, Bonferroni-adjusted)
X1	10.637	0.0049	0.320	$0 \neq 2$, $1 \neq 2$
X2	18.541	0.0001	0.613	$0 \neq 1$, $0 \neq 2$, $1 \neq 2$
X3	9.327	0.0094	0.271	$0 \neq 2$, $1 \neq 2$

As shown in Table 4, all three indicators yielded statistically significant differences across clusters ($p < 0.01$), with large effect sizes, particularly for X2 ($\varepsilon^2 = 0.613$), indicating strong variation in student-teacher ratios among clusters. Post-hoc Dunn tests further confirmed that Cluster 2 consistently differed from the other two clusters across all indicators, reinforcing its label as a high-risk region with elevated dropout rates and resource constraints. The inclusion of effect sizes (ε^2) also demonstrates that the differences between clusters are not only statistically significant, but practically meaningful, supporting the robustness and interpretability of the clustering results in guiding policy decisions. While the three indicators used in the Kruskal-Wallis test are the same variables used in the clustering process, the purpose of this analysis is not to re-discover the clusters, but to provide a formal statistical validation of the observed group differences. This step complements the visual and spatial interpretation by offering quantitative confirmation that the clusters indeed differ significantly in terms of the underlying educational indicators. The use of effect sizes (ε^2) further supports the practical significance of these differences, reinforcing the robustness and interpretability of the clustering outcome. Although this analysis uses the same variables as those used for clustering, the test was conducted to statistically confirm the distinctiveness of the resulting clusters. Rather than being redundant, it serves as an internal consistency check, supporting the interpretation of clusters as meaningful groupings for policy analysis.

The application of the quartile method and the K-Means algorithm produced complementary yet distinct insights into regional educational disparities in West Java. The quartile approach, applied separately to each indicator, highlighted extremes in student-school ratios, student-teacher ratios, and dropout rates, but could not capture interrelations among them. In contrast, the multivariate K-Means clustering integrated all indicators and identified distinct groups that reveal how high density and dropout risks interact across regions. While the quartile method offers simplicity and clarity for indicator-specific targeting, K-Means provides a holistic view of multidimensional regional profiles. Together, these methods show that educational inequality in West Java is influenced by both infrastructural and socio-economic factors, emphasizing the need to combine univariate and multivariate analyses for comprehensive policy insights.

3.2. Discussion

Building on the clustering results and their spatial patterns, this section proposes targeted policy interventions for each group of regions. By aligning strategies with the unique characteristics of each cluster such as infrastructure availability, student-teacher ratios, and dropout risks, policymakers can design more effective and equitable education programs.

Cluster 1, categorized as high-density but stable, includes urban and peri-urban regions. The main challenge in these areas lies in structural pressure, reflected in high student-school and student-teacher ratios. However, the relatively low dropout rates suggest that the system, though overloaded, remains functional. In response, the study proposes establishing regional teacher hubs to better prepare educators for high-pressure settings and piloting a teacher fatigue monitoring system to inform incentive or voluntary relocation policies. Policy responses should prioritize infrastructure expansion and teacher redistribution, in line with the Ministry of Education’s teacher allocation policies (Permendikdasmen No. 1/2025). Previous studies have emphasized that improving teacher distribution in dense urban contexts significantly reduces inequality in learning outcomes[5]. However, uneven teacher distribution has been attributed to the absence of strong legal frameworks, weak education data systems, poor enforcement mechanisms, and local political elite interference, all of which contribute to disparities in education quality, low graduation rates, social and economic inequality, and decreased student motivation and academic achievement[14][15].

Cluster 2, labeled elevated dropout risk that exhibit relatively lower structural density but significantly higher dropout rates. This pattern indicates that socio-economic factors, rather than infrastructure shortages, constitute the main barrier to educational participation. To address this, the study recommends implementing risk-based scholarship schemes targeting students from highly disadvantaged backgrounds, as well as school-based reintegration programs for dropouts involving home visits, counseling, and collaboration with social workers. Thus, interventions should focus on social support programs, including scholarships and conditional cash transfers like PIP (Program Indonesia Pintar), as well as targeted initiatives like School Operational Assistance (BOS): BOS Afirmasi and BOS Kinerja (Permendikbudristek No. 16/2021, No. 24/2020). These findings are consistent with earlier research showing that economic vulnerability is strongly correlated with school dropout in rural and semi-rural regions[16]. Budget allocation for dropout prevention remains necessary, as dropout trends have declined since funding was specifically directed toward targeted intervention programs[17]. Prior studies employing K-Means clustering in dropout risk analysis demonstrate its utility in distinguishing student groups by academic and socio-economic conditions[18]. These findings reinforce the importance of continued education reform to address persistent urban-rural and socio-economic gaps.

Cluster 0, identified as balanced and stable, includes areas with relatively proportional ratios and low dropout rates. Although these regions do not face urgent structural or socio-economic pressures, continuous monitoring is essential to prevent emerging disparities. The study recommends developing an early-warning dashboard using integrated education data systems to detect subtle negative trends such as declining teacher ratios or gradual increases in dropout rates. Ongoing quality assurance, teacher training, and monitoring are necessary to prevent future disparities. This aligns with national education strategies that emphasize sustaining equity gains through continuous professional development and localized monitoring systems such as Dapodik (Data Pokok Pendidikan). Previous research on digital education highlights that reliable infrastructure and targeted teacher training are key determinants of student engagement and the effective integration of educational technologies in rural schools[19]. Improved policy formulation and implementation, including enhanced teacher training, technology integration, and equitable resource allocation, are essential to overcoming current challenges and achieving a resilient, inclusive education system[20]. Educational authorities can leverage clustering results to strategically address teacher shortages by reallocating teaching resources more effectively, thereby potentially improving educational outcomes in underserved areas. To translate these findings into actionable policy, a cluster-based intervention matrix is proposed in table 5, detailing specific strategies, targeted regions, and measurable monitoring indicators.

Table 5. Policy recommendations by cluster.

Cluster	Main Issues	Policy Recommendations	Indicators and Targets
0 (Balanced and Stable)	Relatively balanced educational resources with low dropout risk	- Maintain existing resource allocation. - Regular monitoring of ratios and dropout trends. - Teacher upskilling and retention programs.	- X1 ≈ 300 - X2 < 20 - X3 < 0.05%
1 (High-Density but Stable)	High student density (X1), moderate teacher burden (X2), but low dropout (X3)	- Expand classroom capacity. - Recruit new teachers in high-density schools. - Redistribute teachers from overstaffed to understaffed areas.	- X1 < 350 - X2 < 20 - Maintain X3 < 0.05%
2 (Elevated Dropout Risk)	Low school density (X1), high teacher burden (X2), and highest dropout (X3)	- Risk-based scholarship schemes. - Re-enrolment and community outreach programs. - Conditional Cash Transfer (CCT) programs like *Program Indonesia Pintar* (PIP). - Targeted teacher placement to rural/disadvantaged areas.	- X3 < 0.05% within 2 years - X2 < 20 - X1 ≥ 270

Overall, the cluster-based approach offers a nuanced understanding of regional disparities in educational challenges, providing data-driven insights for targeted planning and resource allocation. By linking recommendations to cluster characteristics and existing regulations, this analysis supports more adaptive and equitable education policies. Although this study does not introduce a new clustering method, it demonstrates the applied use of standard K-Means integrated with spatial analysis to identify sub-provincial patterns of educational inequality in Indonesia, offering practical guidance for early interventions and policy decisions, especially in resource-limited regions.

3.3. Limitations and Future Research

This study has several limitations that warrant consideration. The analysis relied on a single year of data (2024) and focused solely on junior secondary schools, limiting its ability to capture longitudinal trends or provide insights across different education levels. Moreover, only three indicators, student-school ratio, student-teacher ratio, and dropout rate, were examined using the K-Means clustering algorithm, chosen for its simplicity and interpretability. Future research should incorporate additional variables such as teacher qualifications, school infrastructure, and socio-economic factors to provide a more comprehensive view of educational disparities, as well as adopt longitudinal data to evaluate policy impacts over time. Exploring alternative or ensemble clustering methods could enhance cluster validity, while extending the analysis to other provinces would enable comparative studies and test the scalability of policy interventions. Although this study does not introduce methodological innovations, it underscores the practical value of applying clustering and spatial analysis to inform education policy, while future work may contribute theoretically by developing hybrid or machine learning–based policy evaluation models.

4. Conclusion

This study demonstrates that clustering methods can effectively uncover disparities in junior secondary education across West Java by analyzing student-school ratios, student-teacher ratios, and dropout rates. The quartile method provided a clear and simple way to identify regions with extreme values in individual indicators, highlighting priority areas for each variable. Meanwhile, K-Means clustering offered a more holistic perspective by grouping regions based on multidimensional similarities, revealing distinct clusters such as Balanced and Stable, High-Density but Stable, and Elevated Dropout Risk. The comparative results indicate that quartile analysis is useful for diagnostic monitoring, while K-Means provides a stronger foundation for integrated policy targeting. The findings suggest that policy responses must be differentiated according to cluster characteristics. These findings confirm that addressing educational inequality requires not only infrastructure expansion in high-density areas but also social and financial support in dropout-prone regions, ensuring more targeted and equitable policies. The research contributes to the development of evidence-based decision-making in education planning by showing how data-driven clustering approaches can align interventions with local contexts. Although the analysis was limited to three indicators and a single year of data, its results emphasize the value of integrating multiple perspectives in monitoring and policy design. Future studies should expand the range of indicators and incorporate longitudinal data, additional socio-economic variables, or alternative clustering techniques to capture dynamic trends more comprehensively and to strengthen the dynamic understanding of equity in Indonesia's education system.

Acknowledgement

The author would like to express sincere gratitude to the Centre for Data and Information Technology (Pusat Data dan Teknologi Informasi), Ministry of Primary and Secondary Education, for providing essential data, research funding, and technical support throughout the study. The assistance, resources, and guidance offered by the Centre have been invaluable in facilitating the completion of this research and the preparation of this paper.

References

[1] D. Dusalan, "13-Year Compulsory Education: A Strategic Step Towards Equitable Access to Children's Education in Indonesia," *J. Islam. Elem. Educ.*, vol. 3, no. 1, pp. 253–263, 2025, doi: 10.32806/islamentary.v3i1.805.

[2] C. O. Muchtar, A. A. Purba, J. Sintya, M. F. Siagian, and N. M. Harahap, "Systematic Literature Review : Examining Indonesia ' s Educational Inequality Factors and Government Equity Policies," *Int. J. Educ. Pract. Policy*, vol. 3, no. 1, pp. 8–16, 2025, doi: 10.61220/ijepp.v3i1.0257.

[3] L. Judijanto, "Challenges and Opportunities in Education Equity through the 13-Year Compulsory Education Program in Indonesia," *Eastasouth J. Learn. Educ.*, vol. 3, no. 01 SE-Articles, pp. 1–8, Mar. 2025, doi: 10.58812/esle.v3i01.510.

[4] W. Winardi, "Decentralization of Education in Indonesia—A Study on Education Development Gaps in the Provincial Areas," *Int. Educ. Stud.*, vol. 10, no. 7, p. 79, 2017, doi: 10.5539/ies.v10n7p79.

[5] and W. K. Setyosari, Punaji, "Teachers Quality and Educational Equality Achievements in Indonesia," *Int. J. Instr.*, vol. 14, no. 2, pp. 811–830, 2021.

[6] J. S. M. García and M. A. Giovine, "Equity and Education: Philosophies and Measurement," *Soc. Indic. Res.*, vol. 178, no. 2, pp. 707–723, 2025, doi: 10.1007/s11205-025-03572-3.

[7] V. S. Fatmawaty, I. Riadi, and H. Herman, "Higher Education Institution Clustering Based on Key Performance Indicators using Quartile Binning Method," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 24, no. 1, pp. 141–154, 2024, doi: 10.30812/matrik.v24i1.4244.

[8] T. Le Quy, G. Friege, and E. Ntoutsi, *A Review of Clustering Models in Educational Data Science Towards Fairness-Aware Learning*. 2023.

[9] M. Qori, "Pengelompokkan Wilayah Berdasarkan Rasio Guru-Murid Pada Jenjang Pendidikan Menggunakan Algoritma K-Means," *J. Inform. Dev.*, vol. 1, no. 1, p. 34, 2022, [Online]. Available:

https://ejournal.itbwigalumajang.ac.id/index.php/jid/article/view/898

[10] Y. Hasnataeni, M. Nurhambali, R. Ardhani, S. Hafsah, and A. Soleh, "Comparison of clustering analysis of K-means, K-medoids, and fuzzy C-means methods: case study of school accreditation in west java," *J. Soft Comput. Explor.*, vol. 6, pp. 79–88, Jun. 2025, doi: 10.52465/joscex.v6i2.575.

[11] Z. Mustakim and R. Kamal, "K-Means Clustering for Classifying the Quality Management of Secondary Education in Indonesia," *Cakrawala Pendidik.*, vol. 40, no. 3, pp. 725–737, 2021, doi: 10.21831/cp.v40i3.40150.

[12] M. Hammam and T. Utomo, "Clustering Data Siswa Putus Sekolah dengan Algoritma K-Means dan DBSCAN," *J. Komput. dan Inform.*, vol. 18, no. 2, pp. 150–159, 2023.

[13] A. Khoirunnisa', F. A. S. Wibowo, and K. Kismiantini, "Perbandingan Analisis Agglomerative Hierarchical Clustering Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat," *Pros. Semin. Pendidik. Mat. dan Mat.*, vol. 7, no. March, 2023, doi: 10.21831/pspmm.v7i1.273.

[14] N. Innayatun and A. Wibowo, "The Impact of Unequal Distribution of Teachers on The Quality of Education in Indonesia," *J. Ilm. Pendidik. Dasar*, vol. 09, no. 03, pp. 4–6, 2024.

[15] J. M. Sidauruk, M. Susilowati, and K. K. Akbar, "Indonesia's Struggle with Education Inequality: Is Reform the Answer?," *Indones. Discourse*, vol. 2, no. 1, pp. 59–84, 2025.

[16] L. Raspatiningrum, S. Rubai'ah, R. S. Nugraha, R. Shodikin, and R. L. Gorni, "Teacher Retention in Rural Indonesian Schools: An Interpretative Phenomenological Analysis of Career Disorientation and Commitment," *J. Educ. Teach.*, vol. 6, no. 3, p. 2025, 2025, doi: 10.51454/jet.v6i3.641.

[17] D. Dahliana and N. Huda, "Does Education Budget Influence School Dropout?," *KnE Soc. Sci.*, vol. 2024, pp. 65–72, 2024, doi: 10.18502/kss.v9i19.16477.

[18] K. Muttaqin, R. Nurhidayah, N. Novianda, A. Ihsan, J. Sultan, and F. Rifqiyah, "Implementation of K-Means Clustering in Mapping Teacher Distribution Using Geographic Information System," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 9, no. 1, pp. 187–196, 2024, doi: 10.21831/elinvo.v9i1.76884.

[19] Y. E. Lestari, Y. Alif Pudin, and V. M. Wibowo, "The Impact of Digital Learning Policies on Educational Equity in Rural Indonesian Schools," *Int. J. Educ. Eval. Policy Anal.*, vol. 1, no. 2, pp. 13–19, 2024.

[20] I. Widiastuti, "Assessing the Impact of Education Policies in Indonesia: Challenges, Achievement, and Future Direction," *AL-ISHLAH J. Pendidik.*, vol. 17, no. 2, pp. 1955–1964, 2025, doi: 10.35445/alishlah.v17i2.6803.

Spatial Analysis of Illegal Economic Activity Risk Zones in Berau Regency Based on VIIRS Nighttime Imagery and Data Mining

N M A Putri[1]*, A Nurcahya[1], M H Nurodin[1], N H F Yasmin[1], A Fadhilah[1], S Himayah[1]

[1]Geographical Information Science Study Program, Universitas Pendidikan Indonesia, Bandung, IndonesiaSAIG23

*Corresponding author’s email: nadyamarssyanda@upi.edu

Abstract.

Berau Regency in East Kalimantan, while rich in natural resources, faces growing threats from illegal mining, illegal logging, and unauthorized trade that often occur in remote areas and remain undetected by conventional monitoring. These activities have led to environmental degradation, financial losses, and ineffective enforcement, highlighting the urgent need for a spatially based monitoring system. This study develops a geospatial framework that employs Visible Infrared Imaging Radiometer Suite (VIIRS) Nighttime Light data as a proxy for human activity to identify potential hotspots of illegal economic activities. By integrating VIIRS Day/Night Band (DNB) imagery with Geographic Information System and spatial data, a rule- based classification was applied to delineate high-risk zones based on thresholds of nighttime light intensity and proximity to conservation areas, settlements, roads, and mining sites. The results demonstrate the capability of remote sensing and geospatial analysis to systematically detect and map areas vulnerable to illegal economic activity, offering practical support for local authorities to improve monitoring, strengthen enforcement, and develop more targeted policy interventions.

Keyword: Ilegal, mining, logging Berau, remote sensing.

1. Introduction

Berau Regency, located in East Kalimantan, is rich in natural resources, including coal, timber, and marine products [1]. However, this economic potential has also triggered the rise of illegal economic activities such as illegal mining, illegal logging, and unauthorized trade [2]. In practice, these activities are not small in scale, field operations by local police and Berau Coal reported the seizure of more than 10 heavy equipment units used for unauthorized mining between 2023 and 2025 [3], along with evidence of illegal excavation pits spread across concession areas [4]. In addition, illegal mining cases in Kelay District alone resulted in the arrest of 9 suspects and the confiscation of hundreds of ulin wood logs [5], while civil society organizations such as Indonesia’s Mining Advocacy Network (JATAM) have identified at least 11 illegal mining sites and more than 123 abandoned mine pits contributing to flooding events in Berau [6]. These activities are difficult to monitor, as they often occur in remote areas and remain undetected by conventional surveillance systems [7]. As a result, local governments suffer revenue losses, while the environment and local communities experience direct negative impacts [8]. Furthermore, there is currently no effective system in place to spatially identify and map areas at risk of illegal economic activity, leading to misdirected enforcement efforts and ineffective policy interventions[9]. Traditional monitoring systems have proven inadequate in anticipating the spatial dynamics of these activities in a comprehensive and timely manner [10]. Therefore, a remote sensing and spatial analysis- based approach is urgently needed to accurately identify high risk zones [11].

Previous studies have shown that VIIRS Nighttime Lights data is very useful for monitoring illegal economic activity at night. For example, VIIRS-DNB radiance has been utilized to uncover the expansion of small-scale gold mining camps in Gorontalo, where nighttime light intensity in the area surged nearly fourfold between 2014 and 2019 [12]. Furthermore, VIIRS-DNB data have been applied to map nighttime illegal, unreported, and unregulated (IUU) fishing hotspots in the South China Sea [10], thereby generating a spatial database to support the updating of maritime law enforcement policies. Integration of the Automatic Identification System (AIS) data with VIIRS-based Vessel Detection (VBD) has also been used to map the intensity of nighttime trawler operations, revealing unrecorded vessels in AIS and bright light sources potentially engaged in illegal fishing. [13]. An integrated system combining VIIRS-DNB imagery and AIS information has even been developed to monitor multi-species fishing vessels with up to 90% accuracy [1], thereby affirming the practical value of VIIRS data in supporting fisheries conservation and enforcement of no-take zones. However, most of these studies primarily focused on the use of nighttime light data and satellite imagery alone, without integrating other relevant datasets and analytical techniques such as data mining for pattern recognition, Geographic Information System (GIS)-based spatial proximity analysis (e.g., creating buffer zones around roads and settlements), the distribution of legally registered businesses, or Land Surface Temperature (LST) data to improve detection accuracy and support comprehensive risk assessment.

Geospatial technology plays a crucial role in detecting and analyzing patterns of illegal economic activities [7]. The use of VIIRS imagery enables the acquisition of nighttime light data, which serves as a proxy for human activity in areas that are often beyond the reach of direct monitoring [11]. By integrating data mining techniques with GIS, spatial analyses can be conducted to identify lighting anomalies potentially associated with illicit activities [9]. Furthermore, a rule-based classification approach can be applied to identify high-risk zones by integrating nighttime light intensity patterns with proximity to critical infrastructure or protected areas [10]. Integrating official business registration data from government agencies such as the Investment and One-Stop Integrated Services Office (DPMPTSP), the Central Bureau of Statistics (BPS), or the Online Single Submission (OSS) system with VIIRS nighttime light imagery enables the identification of "spatial anomalies," namely locations exhibiting significant economic activity (e.g., strong nighttime illumination) that are not listed in official records [16].

This study aims to identify patterns of illegal economic activity in Berau Regency by utilizing nighttime images from the VIIRS sensor as indicators of human activity in remote areas. A data mining-based spatial analysis model is developed to detect zones at risk of illegal economic activities, employing a rule-based classification technique. Furthermore, a risk zone map is generated as a decision-support tool for local governments and law enforcement agencies. The research also evaluates the correlation between nighttime light intensity and the presence of illegal activities, particularly in areas adjacent to conservation zones or critical infrastructure. The novelty of this study lies in providing an efficient and replicable geospatial technology-based approach to monitor and support law enforcement efforts against illegal economic activities in geographically challenging regions.

2. Research Method

2.1. *Study Area and Period*

The study area is Berau Regency, East Kalimantan, covering diverse land cover from forests to mining sites. The analysis period was January-December 2024.

2.2. *Data Sources*

The data used in this study consists of spatial and non-spatial data, as summarized in Table 1.

Table 1. Research Data Sources and Types

	Type	Source
Administrative boundaries (Kabupaten Berau, Kalimantan Timur, Kalimantan Utara)	Shapefile (Polygon)	tanahair.indonesia.go.id
Land Cover	Shapefile (Polygon)	tanahair.indonesia.go.id
Roads (arteri & kolektor	Shapefile (Polyline)	tanahair.indonesia.go.id
VIIRS NTL Imagery	Raster (GeoTIFF)	Google Earth Engine
Legal businesses	CSV (Point data)	Scraped from Google Maps

Data collection integrated multiple spatial and non-spatial sources to support the spatial analysis of illegal economic activity risk zones in Berau Regency. Administrative boundaries for Berau Regency, East Kalimantan, and North Kalimantan were obtained from the official government portal (tanahair.indonesia.go.id) to delimit the study area. Land cover data, including forests, settlements, and mining areas, were used to assess overlaps with potential illegal activities. Road network data, comprising arterial and collector roads, facilitated the analysis of illuminated zones' accessibility. Nighttime light imagery from the VIIRS-DNB dataset provided by the National Oceanic and Atmospheric Administration (NOAA) specifically the monthly cloud-free composite (version VCMCFG) for 2024 was downloaded as 500-meter resolution GeoTIFF rasters via Google Earth Engine, serving as an indicator of nocturnal economic activity. Legal business locations collected from Google Maps via web scraping provided a reference to differentiate between legal and potentially illegal economic zones. Imaging Radiometer Suite Day/Night Band (VIIRS-DNB) Monthly Cloud-Free Composites (VCMSFG) product, accessed via Google Earth Engine (dataset code: NOAA/VIIRS/DNB/MONTHLY_V1/VCMSFG). Monthly data from January to December 2024 were aggregated, and the mean annual radiance was computed to minimize temporal anomalies such as cloud cover, lunar illumination effects, and short-term non-economic lighting. This approach aligns with the methodology of Kimijima et al. (2021), who applied VIIRS data to detect small-scale gold mining, and Gibson (2025), who validated the use of nighttime light intensity as a reliable proxy for regional economic activity.

2.3. *Data AnalysisWorflow*

The research workflow integrates multiple spatial datasets and geospatial analyses, starting from preprocessing, spatial scoring, to wighted classification. The complete workflow can be seen in Figure 1.

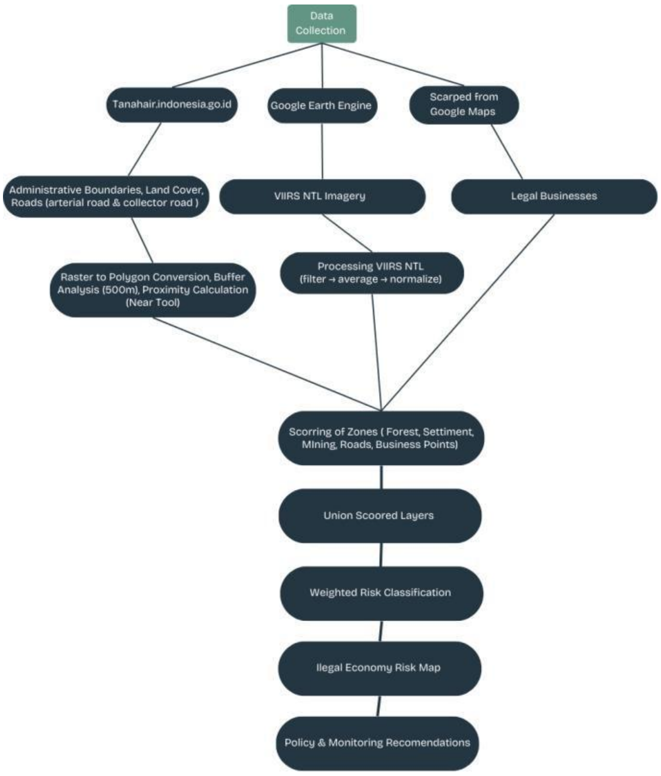

Data Collection
Tanahair.indonesia.go.id
Google Earth Engine
Scarped from Google Maps
Administrative Boundaries, Land Cover, Roads (arterial road & collector road)
VIIRS NTL Imagery
Legal Businesses
Raster to Polygon Conversion, Buffer Analysis (500m), Proximity Calculation (Near Tool)
Processing VIIRS NTL (filter → average → normalize)
Scorring of Zones (Forest, Settiment, MIning, Roads, Business Points)
Union Scoored Layers
Weighted Risk Classification
Ilegal Economy Risk Map
Policy & Monitoring Recomendations

Figure 1. Research flowchart

The workflow for identifying illegal economy risk zones integrates multiple spatial datasets and geospatial analyses. Initially, data were collected, including administrative boundaries, land cover, road networks, VIIRS nighttime light imagery, and existing business locations. The VIIRS data were then preprocessed through filtering, temporal averaging, and normalization to quantify economic activity intensity. The processed raster data were converted into polygons to enable vector-based spatial analyses. Buffer analyses (500 m) were applied to key features, and proximity calculations were conducted using the Near Tool to assess spatial relationships. Each polygon was subsequently scored based on forest cover, settlement density, mining activity, road proximity, and business presence. Scored layers were then overlaid and integrated through a union operation, followed by weighted classification to produce a comprehensive risk assessment. The resulting map delineates areas with varying levels of vulnerability to illegal economic activities and provides a spatial basis for policy formulation and targeted monitoring efforts.

Preliminary analysis was conducted using VIIRS nighttime light data for Berau Regency from January to December 2024. Monthly measurements were averaged to generate an annual mean and subsequently normalized to a 0–1 scale to facilitate weighting in spatial analyses. The normalized raster data were converted into vector polygons to improve spatial precision and enable vector-based operations. Buffers of 500 meters were created around collector roads, arterial roads, administrative boundaries, and legal business locations to delineate zones of influence for accessibility and connectivity assessment. Spatial proximity analysis was performed using the Near tool to compute the shortest distance from each feature to nighttime light sources and legal business points within the 500-meter radius, providing indicators of potential economic activity. In parallel, spatial data mining techniques were applied to extract patterns from both spatial and non-spatial datasets, while legal business location data were collected through automated web scraping of Google Maps to ensure up-to-date coverage. A rule-based classification was implemented to assign binary scores to areas such as forests, settlements, mining sites, collector roads, and arterial roads, where a value of 1 indicated presence within 500 meters of a light source or legal business point and 0 indicated absence. These scored layers were integrated through a union operation to produce a composite layer consolidating spatial information for subsequent analyses. Finally, a

proximity-weighted scoring system was applied to assign relative weights to the identified zones, reflecting their susceptibility to illegal economic activities and guiding prioritization for monitoring and intervention.

A rule-based classification system was developed to assess the risk of illegal economic activity. The classication scheme is presented in Table 2.

Table 2. Classification Score of Night Light Conditions

Condition	Classification	Score	Source
Night Light near Forest and Far From The Roads, Settlements, Legal Business Points	High Risk	0,9	Schlutow & Schöder (2021)
Night Light not Appeararing too High, but Close to the Forest	Moderately High Risk	0,7	Chen et al. (2020)
Night Light near Mining Area	Moderate Risk	0,6	Kimijima et al. (2021)
Night Light Near Roads, Settlements, and Legal Business Points	Not Risk	0,2	Gibson (2025)
Additionally/Other Combinations	Low Risk	Automatically Calculate	

The vulnerability classification was developed through a **rule-based spatial analysis** integrating multiple geospatial indicators, including nighttime light intensity (VIIRS), proximity to forest areas, settlements, road networks, and legal business points. This method enables the identification of zones that are more susceptible to illegal economic activities based on their spatial characteristics and environmental context. The vulnerability levels were determined by combining both spatial proximity and radiance intensity values. Higher vulnerability is associated with remote areas exhibiting persistent nighttime illumination, particularly in or near forested or mining zones, whereas lower vulnerability is found in well-connected and brightly lit urban regions. Nighttime light intensity serves as a spatial proxy for human and economic activities. Regions exhibiting high radiance but located far from formal settlements or infrastructure are indicative of unregulated economic operations, such as illegal mining or logging. This relationship aligns with findings by the World Bank (2020), which emphasize the strong correlation between nighttime luminosity and local economic performance.

The classification of illegal economic risk zones was conducted using a **rule-based spatial classification method** integrated within a Geographic Information System (GIS) environment. This approach combines multiple spatial criteria to identify areas with a higher likelihood of unregulated or illicit economic activities based on geospatial relationships and environmental indicators. The analysis utilitized four primary parameters:

1. Nighttime Light Intensity (VIIRS): Representing the level of nocturnal human activity, with higher values suggesting more active or industrial zones.
2. Proximity to Roads: distance buffers (500 m) from arterial and collector roads were created to assess accessibility and potential logistical support for economic activities.
3. Land Cover: focusing on forest, settlement, and mining areas to identify environmental contexts that may influence illegal activity.

4. Proximity to Legal Business Point: used to distinguish between regulated and unregulated economic clusters.

Each variable was assigned a **weight** according to its relative contribution to illegal economic potential:

1. VIIRS = 0.4
2. Roads = 0.3
3. Forest Area = 0.2
4. Settlements = 0.1

The final composite risk score for each spatial unit was derived using a weighted overlay process, with the following classification thresholds:

1. High Risk: total score > 0.6
2. Medium Risk: total score between 0.3 - 0.6
3. Low Risk: total score 0.3

This method aligns with the **Multi-Criteria Decision Analysis (MCDA)** framework, which integrates both environmental and socio-economic spatial indicators to delineate risk zones (Pedersen et al., 2023). The approach also reflects the principles of **rule-based geographic modeling**, emphasizing spatial relationships between human activity and land characteristics in identifying potential zones of illegal economic concentration.

3. Result and Discussion

3.1 Land Surface Temperature (LST)

The 2024 Land Surface temperature (LST) distribution map of Berau Regency reveals significant spatial temparature variation. The details are illustrated in Figure 2.

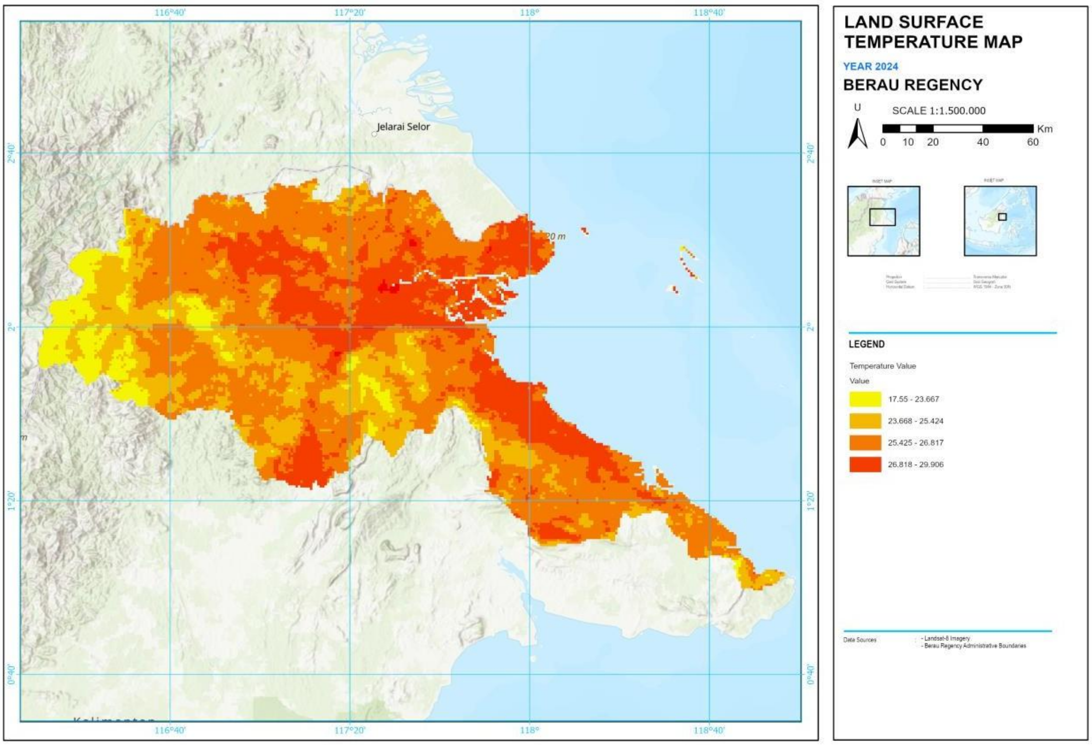


Figure 2. Spatial distribution of Land Surface Temperature (LST)

The 2024 Land Surface Temperature (LST) distribution map of Berau Regency reveals a significant temperature variation, ranging from approximately 17°C to nearly 33°C. The highest surface temperatures are predominantly concentrated in the southern, eastern, and northeastern parts of the regency, particularly around Talisayan, Biatan, Tabalar, and parts of Teluk Bayur districts. These areas are depicted in dark red, indicating very high surface temperatures. In contrast, the western to southwestern regions, including Segah, Kelay, and parts of Gunung Tabur, generally exhibit cooler surface temperatures between 17°C and 23°C, represented by yellow tones. These relatively cooler conditions are likely influenced by the presence of dense vegetation cover, including tropical rainforest in highland areas, which remain relatively undisturbed and distant from major human activities and coastal zones.

The Land Surface Temperature (LST) results indicate that the highest surface temperatures are concentrated in the eastern and southeastern regions of Berau Regency, particularly in the Talisayan and Biatan districts, with temperature ranges between 30°C and 33°C. The increase in LST in these areas correlates strongly with intensive extractive economic activities, such as mining operations and deforestation, which reduce vegetation cover and disrupt the natural surface heat balance. According to Ghulam et al. (2020), variations in LST can serve as a reliable indicator of anthropogenic land-use changes, especially those associated with resource exploitation and industrial expansion. In this context, elevated LST values not only reflect environmental degradation but also indirectly represent intensified economic activities, including illegal or unregulated mining.

This pattern aligns with findings in similar studies (e.g., Kimijima et al., 2021), where increased surface temperatures were detected in areas of artisanal and small-scale mining (ASM). The removal of vegetation and soil layers for mineral extraction increases heat absorption, thereby enhancing the thermal signal observable via satellite data. Thus, LST provides an important complementary variable to nighttime light (VIIRS) analysis in identifying potential zones of illegal or informal economic activity. higher LST values in Berau Regency can be interpreted as a proxy for the spatial intensity of human-induced land transformation, particularly those linked to the expansion of extractive economic sectors. When analyzed together with VIIRS nighttime light data, LST contributes to a more comprehensive understanding of the spatial dynamics of economic vulnerability and informality within the region.

The pronounced temperature differences can be attributed to multiple environmental factors. One of the primary drivers of elevated temperatures in the eastern and southeastern regions is land cover change caused by human activities, particularly large-scale deforestation. In districts such as Talisayan and Biatan, intensive land clearing both for legal and illegal mining has resulted in the loss of natural vegetation, thereby reducing the cooling effect provided by evapotranspiration. This observation is consistent with findings by Ghulam et al. who reported that vegetation loss significantly contributes to surface temperature rise. Furthermore, the geographic characteristics of eastern and southeastern Berau, located in lowland areas and near the coast, increase their susceptibility to warming. This is compounded by higher concentrations of human activity, which further amplifies the urban heat island effect.

An equally critical issue is the proliferation of illegal mining activities in Berau, particularly in eastern areas such as Talisayan and Biatan. According to Kompas, unauthorized mining operations (PETI) persist in numerous locations, resulting in uncontrolled land clearing without environmental reclamation. This not only exacerbates environmental degradation but also accelerates the increase in surface temperatures across these regions. The impacts of rising surface temperatures extend beyond environmental consequences and significantly affect social conditions. These include reduced agricultural productivity, disruption of clean water availability, and heightened health risks related to heat exposure. Over the long term, higher temperatures also increase the likelihood of forest and land fires, particularly during dry seasons. Therefore, the LST map serves not only as a tool for temperature monitoring but also as an essential indicator for assessing environmental degradation risks linked to illegal mining activities in Berau.

3.1. Nighttime Light Intensity

The spatial distribution of nighttime light intensity is displayed in Figure 3.

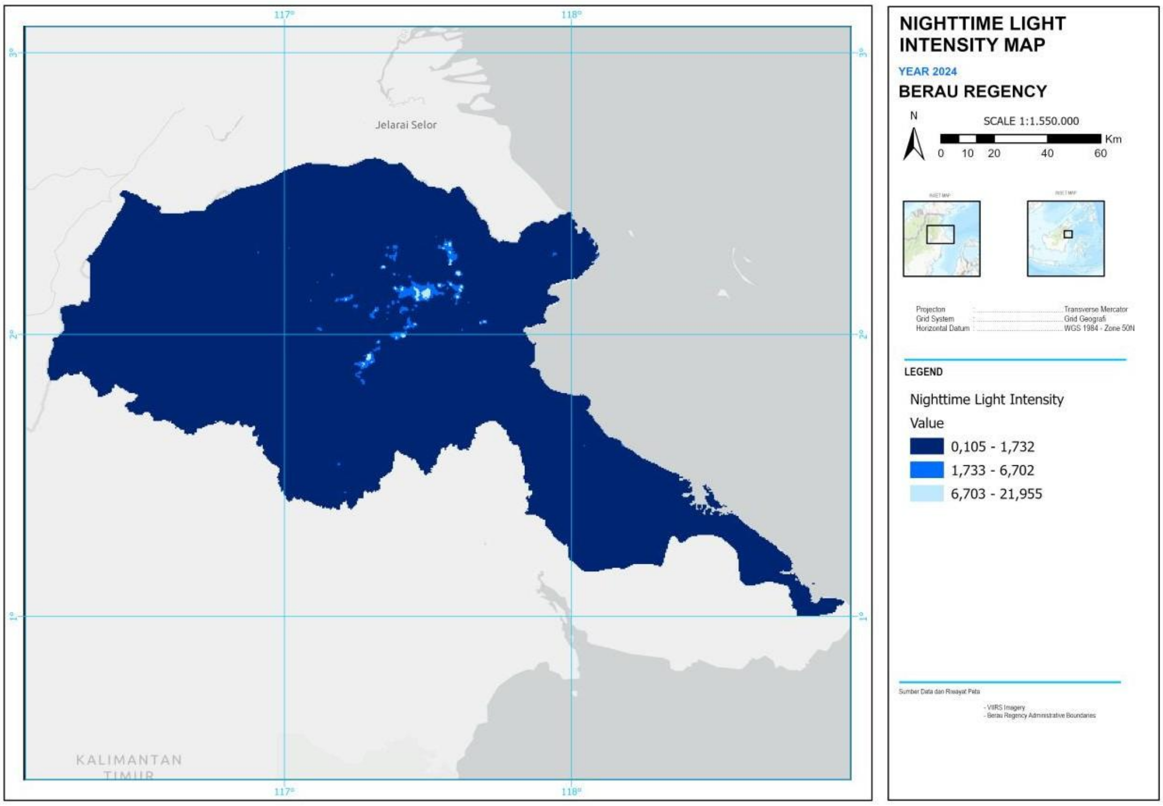

117°
118°
3°
2°
1°
Jelarai Selor
KALIMANTAN TIMUR
NIGHTTIME LIGHT INTENSITY MAP
YEAR 2024
BERAU REGENCY
N
SCALE 1:1.550.000
0 10 20 40 60 Km
INSET MAP
INSET MAP
Projecton : Transverse Mercator
Grid System : Grid Geografi
Horizontal Datum : WGS 1984 - Zone 50N
LEGEND
Nighttime Light Intensity
Value
0,105 - 1,732
1,733 - 6,702
6,703 - 21,955
Sumber Data dan Riwayat Peta
- VIIRS Imagery
- Berau Regency Administrative Boundaries

Figure 3. Nighttime Light Intensity Map of Berau

The average nighttime light map of Berau Regency illustrates the spatial distribution of light intensity based on year-long observations from VIIRS imagery in 2024. The classification divides the region into three categories: low (natural or insignificant illumination), medium (small settlements), and high (strong indications of nighttime activity).

Most areas in Berau Regency fall under the low-intensity category, reflecting extensive tropical forest zones, conservation areas, and regions with limited economic or infrastructural development. High-intensity zones are primarily concentrated around Tanjung Redeb and its surroundings, which serve as the administrative and economic hub of the regency. This concentration indicates elevated nighttime activity, often associated with public facilities, dense residential areas, and industries such as mining and services. Light distribution also appears around the main urban core and is scattered across several suburban and small settlement areas, representing moderate economic activity and limited development compared to the central urban zone.

To complement the spatial visualization, the temporal fluctuation of light intensity throughout 2024 is shown in Figure 4.

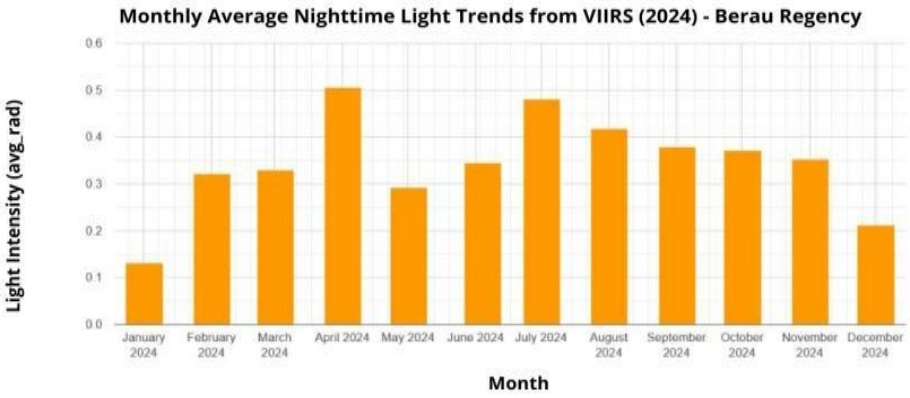


Figure 4. Graph of Average Nighttime Light Intensity Trends (in average radiance units)

The trend graph of average nighttime light intensity (expressed in average radiance) reveals fluctuations throughout 2024. Peak intensities were recorded in April and July, likely corresponding to increased economic activities, production cycles, or industrial operations such as mining and construction projects that were particularly active during these periods. In contrast, the lowest intensity levels occurred in January and December, typically coinciding with the rainy season or extended year-end holidays, which generally lead to reduced nighttime activity.

This variability in nighttime illumination serves as an indirect indicator of local economic dynamics, reflecting patterns such as changes in industrial operations, the development of new residential areas, or the presence of dispersed informal economic activities in non-urban regions. These spatial and temporal patterns of nighttime light intensity are highly relevant for detecting potential hotspots of illegal economic activities, as abnormal or concentrated light emissions in remote or conservation areas may indicate unauthorized mining, logging, or other illicit operations.

3.2. Illegal Economy Risk Zones

The integrated spatial analysis produced a risk zone map for illegal economic activities, as presented in Figure 5.

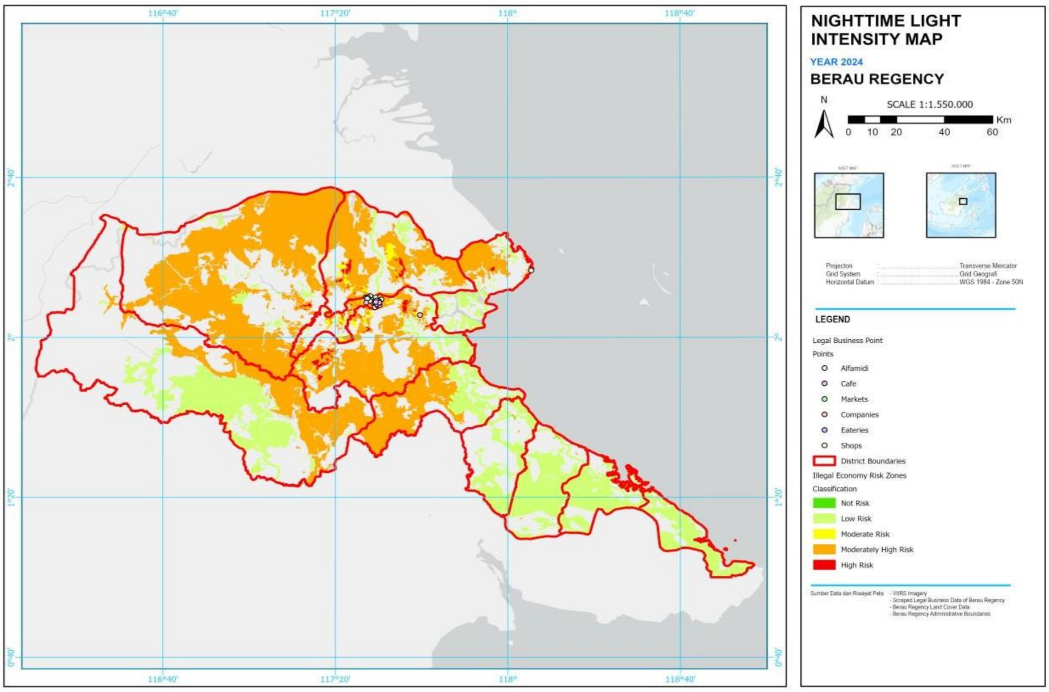

116°40'
117°20'
118°
118°40'
2°40'
2°
1°20'
0°40'
NIGHTTIME LIGHT INTENSITY MAP
YEAR 2024
BERAU REGENCY
N
SCALE 1:1.550.000
0 10 20 40 60 Km
Projecton : Transverse Mercator
Grid System : Grid Geografi
Horizontal Datum : WGS 1984 - Zone 50N
LEGEND
Legal Business Point
Points
Alfamidi
Cafe
Markets
Companies
Eateries
Shops
District Boundaries
Illegal Economy Risk Zones
Classification
Not Risk
Low Risk
Moderate Risk
Moderately High Risk
High Risk
Sumber Data dan Riwayat Peta
- VIIRS Imagery
- Scraped Legal Business Data of Berau Regency
- Berau Regency Land Cover Data
- Berau Regency Administrative Boundaries

Figure 5. Illegal Economy Risk Zones of Berau in 2024

The Illegal Economic Risk Zone Map of Berau Regency for 2024 represents the outcome of a spatial analysis integrating VIIRS (Visible Infrared Imaging Radiometer Suite) nighttime imagery with ancillary data such as legal business locations and spatial elements including road networks, settlements, land cover (particularly forest areas), and mining sites. The primary objective of this mapping is to identify areas with spatial proximity to legal entities and unregistered human activities.

The results indicate that areas classified as moderately high to high risk, depicted in orange and red are predominantly distributed in the central region of Berau Regency, particularly around Tanjung Redeb, Teluk Bayur, and extensive forest zones within Gunung Tabur and Segah Districts. These zones are generally characterized by proximity to forest areas and legal business points, combined with high nighttime light intensity as captured by VIIRS imagery. This pattern suggests potential nighttime activities occurring near registered enterprises, which may include unauthorized operational expansion, unlicensed resource utilization, or shadow economy practices that exploit formal infrastructure without proper documentation. The proximity of risk zones to legal businesses also implies possible overlaps between formal and informal activities, complicating monitoring and law enforcement efforts on the ground.

Tanjung Redeb is recognized as a hub for large- and medium-scale mining operations managed by licensed companies. However, the high intensity of economic activity in this sector often operating within complex and secluded geographic spaces creates opportunities for illegal practices such as unlicensed mining, off-channel trade of mineral products, and covert business activities within corporate perimeters. The spatial proximity between legal enterprises and formal mining zones strongly indicates a dynamic transitional zone where legality and illegality coexist. In other words, the presence of formal companies does not eliminate governance risks; rather, it introduces gray areas that enable illegal economic activities to thrive through permit misuse, infrastructure access exploitation, or the involvement of informal actors in surrounding areas.

The statistical analysis in this study was conducted to interpret the spatial distribution of nighttime light intensity and its relationship with the occurrence of illegal economic activities across Berau Regency. The analysis utilized aggregated mean radiance values from VIIRS imagery, spatial frequency of nighttime activity clusters, and proximity measures to key geographic features such as road networks, forest boundaries, mining concessions, and settlement areas.

The results indicate a significant spatial correlation between areas of persistent nighttime illumination and zones with active or expanding economic operations. In particular, regions displaying high radiance values but located far from major roads and settlements often coincide with forested or mining zones, which are commonly associated with informal or illegal economic practices. Conversely, areas with bright and dense nighttime light signatures near urban centers, such as Tanjung Redeb and Sambaliung, correspond to legal and regulated economic activities.

To validate these findings, comparisons were made with previous studies that have applied VIIRS data for economic and environmental monitoring. Kimijima et al. (2021) identified a strong relationship between increased nighttime brightness and unregulated gold mining activities in Gorontalo, Indonesia. Similarly, Li et al. (2021) demonstrated the successful integration of VIIRS data and Automatic Identification System (AIS) signals to detect illegal vessel operations in the South China Sea. The patterns observed in Berau are consistent with these studies, suggesting that VIIRS nighttime lights can effectively serve as an indirect proxy for detecting and monitoring illicit economic behaviors. The results highlight the potential of integrating remote sensing, spatial statistics, and socio-economic indicators to understand and quantify the dynamics of illegal economic activity. The findings not only align with global research on nighttime light analysis but also provide valuable insights for regional economic governance and spatial planning in resource-dependent regions such as Berau.

From an economic governance perspective, these findings highlight the need for stricter spatial monitoring mechanisms in mining-intensive regions. Furthermore, an integrative approach involving local government, mining companies, and communities is essential to strengthen oversight, ensure supply chain transparency, and enhance social engagement. This map can also serve as a strategic reference for formulating economic zoning and spatial planning policies by incorporating spatial vulnerability to irregular economic activities.

The implications of this study are significant for regional economic development policy and spatial governance. First, the map can guide the prioritization of monitoring efforts for informal economic activities, particularly in areas combining forest adjacency with limited formal infrastructure access. Second, the findings underscore the importance of expanding legal infrastructure networks, such as roads and administrative services, to reduce spatial inequality and strengthen state control in vulnerable zones. Third, the map provides a basis for developing land-use policies and spatial control strategies that adapt to hidden socio-economic risks. Finally, this approach supports the design of long-term programs for legalization, supervision, and empowerment of small and medium enterprises in areas exhibiting illegal economic activity, as part of a strategy to integrate the informal economy into the formal system. This map functions not only as a tool for spatial vulnerability assessment but also as an evidence-based foundation for policy formulation aimed at achieving inclusive and sustainable economic governance in Berau Regency.

4. Conclusion

The analysis reveals that areas with a high potential for illegal economic activities are concentrated near forested zones in Gunung Tabur and Segah Districts, with a broader distribution in the central part of Berau Regency, particularly around Tanjung Redeb and Teluk Bayur. Tanjung Redeb, as the hub of large- and medium-scale mining operations managed by licensed companies, presents a complex and restricted geographic environment that inadvertently creates opportunities for illegal practices to emerge. The spatial proximity between legal enterprises and formal mining sites strongly indicates a dynamic transition zone where legality and illegality intersect. The detected nighttime activity patterns, which are not officially recorded, emphasize the urgency of implementing stricter spatial monitoring systems, fostering collaborative governance among authorities, private companies, and local communities, and adopting data-driven approaches to inform spatial planning and economic zoning policies. These findings underscore the critical role of integrated geospatial analysis in strengthening law enforcement and sustainable resource governance in resource-rich yet vulnerable regions.

Acknowledgement

The author would like to express sincere gratitude to the Study Program of Geographic Information Science, Faculty of Social Science Education, Universitas Pendidikan Indonesia, for academic guidance and continuous support during the research process. Appreciation is also extended to the Berau Regency Government and Berau Coal for providing access to spatial datasets and field information that were essential for data validation. The author also thanks colleagues and peers who contributed constructive feedback, which significantly improved the quality of this manuscript.

References

[1] Badan Pusat Statistik (BPS), “Kabupaten Berau Dalam Angka 2024,” BPS Berau, 2024. [Online].Available: https://beraukab.bps.go.id

[2] CIFOR, “Illegal Logging and Mining in East Kalimantan: A Case Study from Berau and Kutai Timur,” CIFOR Report, 2006. [Online]. Available: https://www.cifor.org/knowledge/publication/1996

[3] PT Jackson Jaya Abadi, “Again, Three Heavy Equipment Seized from Illegal Mine in Berau,” News Release, Jun. 2024. [Online]. Available: https://jacksonjayaabadi.com/en/again-three-heavy-equipment- seized-from-illegal-mine-inberau

[4] Reuters, “Indonesia to launch crackdown on illegal mines in forests,” Reuters, Aug. 28, 2025. [Online]. Available: https://www.reuters.com/business/environment/indonesia-launch-crackdown- illegal-mines-forests-2025-08-28

[5] ANTARA News, “Indonesia to crack down on illegal mining in forest areas,” ANTARA News, Aug. 22, 2025. [Online]. Available: https://en.antaranews.com/news/374921/indonesia-to-crack-down-on- illegal-mining-in-forest-areas

[6] JATAM, “Illegal Mining in East Kalimantan,” JATAM Report, 2024. [Online]. Available:https://jatam.org

[7] J. Gibson, “Lost in Translation? A Critical Review of Economics’ Use of Night Lights,” Remote Sens., vol. 17, no. 7, p. 1130, 2025. doi: 10.3390/rs17071130

[8] Mongabay, “Indonesian campaigns getting money from illegal logging, mining,” Mongabay News, Mar. 21, 2023. [Online]. Available: https://news.mongabay.com/2023/03/indonesian-campaigns- getting-money-from-illegal-logging-mining-watchdog-says

[9] World Bank, “Shedding Light on Night Lights Data: DMSP vs. VIIRS,” World Bank Blogs, Jun. 8, 2020. [Online]. Available: https://blogs.worldbank.org/en/opendata/shedding-light-night-lights-data- dmsp-vs-viirs

[10] A. Schlutow and W. Schröder, "Rule-based classification and mapping of ecosystem services with data on the integrity of forest ecosystems," Environmental Sciences Europe, vol. 33, no. 1, art. no. 50, 2021. [Online]. Available: https://link.springer.com/article/10.1186/s12302-021-00481-3

[11] F.-C. Hsu, C. Elvidge, K. Baugh, and T. Ghosh, “Cross-Matching VIIRS Boat Detections with Vessel Monitoring Data in Indonesia,” Remote Sens., vol. 11, no. 9, p. 995, 2019. doi: 10.3390/rs11090995

[12] S. Kimijima, T. Sano, K. Kubo, and N. Yamaguchi, “Detection of Artisanal and Small-Scale Gold Mining Activities Using Nighttime Light and Precipitation Data,” Int. J. Environ. Res. Public Health, vol. 18, no. 20, p. 10786, 2021. doi: 10.3390/ijerph182010786

[13] R. Li, J. Xu, and Z. Chen, “Integrating AIS and VIIRS-Based Vessel Detection for Maritime Surveillance,” Fish. Res., vol. 243, p. 106070, 2021. doi: 10.1016/j.fishres.2021.106070

[14] H. Tian, J. Zhao, and Y. Sun, “An Integrated System for Monitoring Multi-Species Fishing Vessels Using VIIRS and AIS,” Mar. Policy, vol. 144, p. 105236, 2022. doi: 10.1016/j.marpol.2022.105236

[15] Reccessary, “Indonesia: 47 Companies in Mining, Palm Oil Tied to Illegal Deforestation,” Green Policy News, Mar. 2025. [Online]. Available: https://www.reccessary.com/en/news/indonesia-47- companies-mining-palm-oil-illegal-deforestation

[16] Z. Chen, B. Yu, C. Yang, Y. Zhou, X. Qian, C. Wang, et al., “An extended time-series (2000–2018) of global NPP-VIIRS-like nighttime light data from a cross-sensor calibration,” *Earth Syst. Sci. Data Discuss.*, pp. 1–34, 2020.