Jurnal Masyarakat Informatika, 16(2), 2025
DOI:10.14710/jmasif.16.2.76056
ISSN: 2777-0648

Regional Stability and Dynamics of Rice Production in West Java through Spatiotemporal Clustering

Restu Puji Hidayat
Indonesian Center for Agricultural Socio-Economics and Policy Studies, Bogor, Indonesia
Corresponding author: restupuji@pertanian.go.id

Abstract

The classification of 23 regencies/cities in West Java (2008–2024) was executed using the K-Means algorithm on a dataset spanning five variables: production, harvested area, productivity, population, and agricultural workforce. K-Means was chosen for its efficiency and ease of interpretability when analyzing large-scale multivariate data across time. Optimal cluster determination involved evaluating the Elbow Method, Silhouette Score, and the Davies-Bouldin Index (DBI). Although K=5 was suggested by the Elbow Method, K=6 was selected because it demonstrated a more stable and representative regional separation, supported by the lowest DBI (0.8221) and a relatively high Silhouette Score (0.4531). Cluster boundaries were further validated through PCA and GIS visualization. The analysis revealed precise regional segmentation. Key findings indicate that Indramayu, Karawang, and Subang regencies are stable, high-production centers, suitable for intensification and modernization. Conversely, regions like Bandung and Garut regencies exhibited dynamic cluster shifts driven by urbanization and climate variability. This segmentation has crucial policy implications: stable areas are suitable for intensification, dynamic areas require adaptive risk-mitigation policies, and urban-influenced regions (Bandung, Bekasi, and Depok cities) must focus on diversification and agricultural innovation. Despite the limitations of K-Means' inability to capture complex, non-linear clusters, this research highlights the value of integrating spatiotemporal clustering for policy insights. Future research should incorporate climate and land-use data with advanced clustering methods, such as DBSCAN and HDBSCAN. HDBSCAN is more suitable for modeling clusters with varying densities, and time-series approaches should also be integrated. Overall, these results provide an essential, evidence-based framework for targeted agricultural planning.

Keywords : *Agricultural Policy, Clustering, Rice Production, Spatiotemporal Analysis, West Java*

1 Introduction

Rice is the primary source of food for more than 278.16 million Indonesians [1]. As a strategic commodity, rice is not only the foundation of food security but also plays a crucial role in national and regional economic stability. Java Island contributes approximately 44.47% of the total national rice production, and West Java is one of the provinces with the most significant contribution to the total national production [2].

However, there are significant spatial differences in rice productivity across regions. For example, Indramayu Regency records productivity above 7 tons/ha [3], while regencies in the East Priangan region, such as Tasikmalaya, tend to be below 6 tons/ha [4]. This disparity is not only caused by agroclimatic conditions and irrigation availability [5], but also by demographic factors such as population size [6] and labor force in the agricultural sector [7]. This disparity has a direct impact on the effectiveness of public policies, such as the distribution of subsidized fertilizers, intensification programs, and the distribution of high-quality seeds. In this context, the government requires accurate spatial information to determine region-based intervention priorities. However, most previous studies have tended to focus on quantitative predictive approaches, such as the use of Support Vector Regression (SVR) considering climatic and economic factors to predict rice production [8], and the use of linear regression, random forest, and k-nearest neighbor to predict rice production in West Java,

Received: 16 July 2025 ; Revised: 02 October 2025; Accepted: 10 October 2025; Published: 16 October 2025

which shows a declining trend despite remaining consistent as one of the top three national producers [9]. Efforts to enhance food security and sustainability require cross-sectoral collaboration, including the integration of technology and the adoption of climate-smart agricultural practices. This also emphasizes the strategic role of relevant ministries in promoting sustainable agricultural development [10].

In other studies, clustering approaches have been used to group administrative regions based on similarities in production characteristics, such as productivity and harvest area [11]. At the international level, Abirami et al. [12] demonstrated that the application of the K-Means algorithm is not only practical in grouping rice-producing regions but also contributes to crop yield predictions by considering climate and irrigation variables, thereby generating relevant spatial information to enhance productivity. However, most previous studies have used clustering approaches within a cross-sectional framework, without considering the critical temporal dynamics in agricultural systems. Yet, annual fluctuations due to climate variables, changes in agricultural policies, and demographic dynamics (such as migration or urbanization) have the potential to shift the cluster structure of regions significantly. A spatiotemporal approach that monitors cluster changes over time is necessary to understand the resilience of areas to systemic disturbances.

A study by Wickramasinghe et al. [13] shows that combining time-series data, climate variables, and machine learning can provide deeper insights into agricultural dynamics. On the other hand, analytical approaches such as Principal Component Analysis (PCA) and the K-Means algorithm have proven effective in improving the quality of high-dimensional data clustering, in terms of visualization, computational efficiency, and cluster validity, as demonstrated by metrics such as the Silhouette Score [14][15], the Elbow method [16], and the Davies–Bouldin Index (DBI) [17].

This study aims to cluster regencies/cities in West Java Province based on five leading agricultural indicators: rice production, harvested area, productivity, population, and the number of workers in the agricultural sector, using the K-Means algorithm. Additionally, this study analyzes the spatiotemporal dynamics of clusters during the 2008–2024 period, including the stability of regional positions within clusters and annual transition trends. The next objective is to identify extreme changes, such as cluster shifts of two levels or more, which may indicate structural instability or significant policy influences on the agricultural sector in the region. The novelty of this research lies in its integration of spatial and temporal dimensions simultaneously at the regency/city level, which previous agricultural clustering studies have rarely addressed. This contribution provides new insights into regional disparities, dynamic structural changes, and their policy implications for adaptive agricultural development in West Java.

2 Literature Review

2.1 Factors Affecting Rice Production

The interaction of biophysical and socio-economic factors influences rice production. Biophysical factors, such as climate (temperature, rainfall, and sunlight intensity), soil conditions, and irrigation systems, play a crucial role. Climate change, particularly rising temperatures and increased rainfall variability, has been shown to impact global rice production significantly. Therefore, adaptation strategies are needed, including the development of climate-resilient varieties, the application of precision agriculture technologies, and policy support to ensure the sustainability of production and food security [18][19]. Based on research findings, population size has been shown to influence rice production in Tapanuli Selatan Regency [6]. This suggests that demographic factors, such as population growth, can impact rice production levels, either through increased consumption demands or the availability of labor in the agricultural sector. Suresh et al. [20] demonstrate that labor is the most significant input in rice production in Sri Lanka. At the same time, non-climatic factors, such as migration, access to training, and seed varieties, also impact farmers' technical efficiency, highlighting the importance of increased capacity and farmer involvement in achieving productivity improvements.

A multivariate approach is necessary because various factors are generally interrelated and produce simultaneous effects. Therefore, the use of spatiotemporal panel data encompassing production and demographic variables provides a more comprehensive basis for data-driven spatiotemporal analysis. In a regional context such as West Java, these variables are essential to analyze simultaneously because regional differences can create distinctive spatiotemporal patterns in rice production.

2.2 Clustering Methods in Agricultural Studies

Clustering is an unsupervised learning method that aims to group objects based on feature similarities. In agriculture, the K-Means algorithm has become a popular method due to its ease of implementation and interpretation of results. A study by Kurniawati et al. [11] categorized regencies/cities in Indonesia based on harvest area and productivity, revealing differences in characteristics among regional groups. Abirami et al. [12] applied K-Means to regencies in Tamil Nadu, India. They showed that regional classification based on rice production variables could serve as a basis for more targeted policy decisions, such as fertilizer subsidies or agricultural machinery incentives.

However, most previous studies still rely on a cross-sectional approach and have not systematically explored spatiotemporal dynamics in the context of agriculture. This study aims to fill this gap by combining spatial and temporal dimensions to produce more informative and relevant clustering, providing a basis for data-driven agricultural policy formulation.

2.3 Spatiotemporal Segmentation and Regional Stability

In addition to the static clustering approach, the spatiotemporal dimension provides a more dynamic perspective in observing changes in the characteristics of agricultural areas over time. Zhu et al. [21] proposed a deep learning-based STMA method that utilizes Sentinel-1 time-series imagery for crop mapping, combining multi-level spatiotemporal attention to improve the accuracy of crop phenology mapping in complex agricultural systems. Supriatna et al. [22] conducted a spatiotemporal analysis of rice growth in Karawang Regency using Sentinel-1 imagery. They found that the distribution of phenological phases does not always follow the irrigation system, and that there is an acceleration of harvest during the dry season. Spatiotemporal analysis expands the scope of clustering by considering changes in regional characteristics over time. Cluster stability is an essential indicator in the spatiotemporal approach, reflecting the extent to which a region maintains its agricultural profile over time. Areas with high stability tend to have stable agricultural conditions. In contrast, areas that frequently experience changes in clusters may indicate structural changes or external disturbances, such as policy changes, urbanization pressures, natural disasters, or internal dynamics, like shifts in primary commodities. Monitoring these dynamics is essential for supporting evidence-based agricultural policy formulation that is responsive to change and adaptive to regional challenges.

However, most previous studies remain limited to cross-sectional or static perspectives and do not systematically integrate both spatial and temporal dimensions. This gap highlights the need for studies, such as the present one, that address these limitations and provide more comprehensive and policy-relevant insights. Future studies are recommended to incorporate climate factors, such as rainfall, temperature, and irrigation patterns, potentially utilizing remote sensing data, to refine clustering results further and enhance their applicability for agricultural policy planning.

3 Research Method

3.1 Data, Data Sources, and Study Area

This study covers 23 regencies/cities in West Java Province from 2008 to 2024. Four regions were excluded: Bogor Regency, West Bandung Regency, Pangandaran Regency, and Cimahi City, as they lacked consistent production and harvest area data throughout the period. Data were sourced from the Central Statistics Agency (BPS) and the Ministry of Agriculture, organized in a panel format (combining cross-section and time-series), which enables the simultaneous exploration of spatial and temporal aspects. The five variables used include:

a. Rice production (tons)
b. Harvested area (ha)
c. Productivity (quintals/ha)
d. Population (people)
e. Agricultural workforce (people)

3.2 Data Pre-processing and Missing Value Handling

Missing values were found in the variable for the population working in the agricultural sector for all regencies in 2016, as well as for the city of Cirebon in 2011. To address this, linear interpolation was performed using the values from the previous and subsequent years at the exact location. Linear interpolation is a univariate imputation method that utilizes data before and after the missing point in a time series at the exact location, without relying on data from other places [23]. The formula for linear interpolation is shown in Eq. (1) where (x_1, y_1) and (x_2, y_2) are known pairs of data points, and (x, y) is the point to be estimated.

$$y = y_0 + (x - x_0) \cdot \frac{y_1 - y_0}{x_1 - x_0} \tag{1}$$

3.3 Data Normalization

Before the clustering process, data normalization was performed using MinMax Scaling to equalize the scale between variables. This step is essential because the K-Means algorithm is sensitive to differences in scale between features, and the use of MinMax Scaling has been proven to improve the accuracy of clustering results [24]. The formula used is shown in Eq. (2) where X is the original value, min(X) is minimum value, dan max(X) is maximum value from all data in that variable.

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)}, \tag{2}$$

3.4 Determination of Cluster Number

3.4.1 Silhouette Score

The Silhouette Score is used as an internal evaluation metric to assess the quality of clustering based on cohesion and separation between clusters, as efficiently implemented by Gaido [25] on large-scale datasets. For each data point i, $a(i)$ is calculated as the average distance to members within the same cluster, and $b(i)$ as the average distance to the nearest other cluster. The silhouette coefficient is calculated using the formula in Eq. (3).

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}} \tag{3}$$

$s(i)$ value range from –1 to 1, where a value close to 1 indicates good clustering. The higher the S value, the better the cluster structure formed. The overall Silhouette Score is calculated as the average of all $s(i)$ value as in the formula in Eq. (4).

$$S = \frac{1}{n}\sum_{i=1}^{n} s(i) \tag{4}$$

3.4.2 Elbow Method

The determination of the optimal number of clusters was performed using the Elbow method, which involves observing changes in the within-cluster sum of squares (WCSS) for various values of k. This method aims to identify the “elbow” point, i.e., the value of k at which adding more clusters no longer results in a significant decrease in inertia. Based on the experimental results, the number of clusters determined through this approach was found to improve the efficiency of the clustering process, with 25% fewer iterations required to achieve convergence compared to using other numbers of clusters [16].

3.4.3 David-Bouldin Index (DBI)

DBI is an internal evaluation metric that measures the ratio between intracluster compactness and intercluster separation, where lower values indicate better clusters. Recent studies have shown that DBI is effective for use on high-density datasets and can provide a quantitative measure of the uncertainty in clustering results [26]. A lower DBI value indicates a better cluster structure. Mathematically, the DBI is formulated as follows in Eq. (5).

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{S_i+S_j}{M_{ij}}\right) \tag{5}$$

Where k is the number of clusters, Si is the average distance between each point in the cluster i with its cluster center (showing dispersion or cohesion), and Mij is the distance between the center of cluster i and cluster j (indicating the separation between clusters).

3.5 K-Means Clustering Algorithm

The K-Means algorithm is one of the most commonly used clustering methods due to its simplicity in implementation and computational efficiency. This algorithm works iteratively to minimize the squared distance between cluster members and their centroids, as first formulated by Hartigan and Wong [27]. The objective function (loss function) of K-Means is defined in Eq. (6).

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} |x - \mu_i|^2 \tag{6}$$

where k is the number of clusters, Ci is the data set in cluster i, x is data point, and μ_i is centroid in cluster i. This function calculates the total square distance between each data point and its cluster centroid, so that the smaller the J value, the better the separation and compactness of the clusters formed.

3.6 Visualization of Clustering Results

Visualization of clustering results is essential for understanding the distribution and structure of groups in data, especially after dimension reduction. Several studies have shown that integrating the K-Means algorithm with techniques such as PCA and cluster-based feature selection can help simplify visualization, reduce computational complexity, and maintain the accuracy of clustering results [14][15]. Additionally, Geographic Information System (GIS) based visualization was employed to map the cluster results across West Java's administrative boundaries. This approach facilitates the identification of spatial patterns, regional similarities, and more contextual policy implications.

3.7 Research Flow

The research flow is illustrated in Figure 1 as follows.

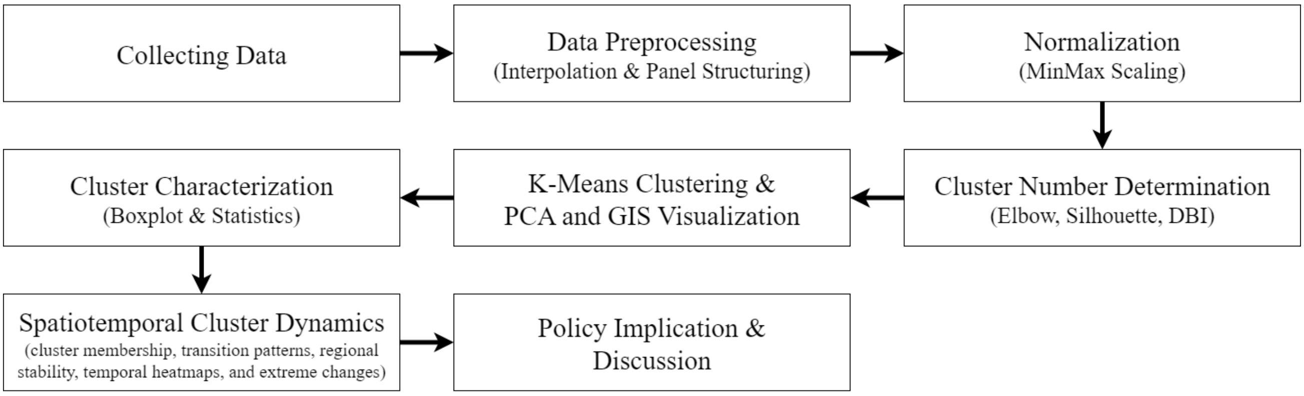


Figure 1. Research flow of research

3.8 Programming Environment

The entire data processing, visualization, and cluster analysis were conducted in Python using Google Colaboratory. Pandas and NumPy were used for data manipulation, while Matplotlib and Seaborn were employed for visualization. scikit-learn was utilized for clustering, normalization (MinMaxScaler), dimensionality reduction (PCA), and evaluation (Elbow Method, Silhouette Score, DBI). Additionally, GeoPandas was used for GIS-based spatial visualization.

4 Results and Discussion

4.1 General Data Description

Figure 2 illustrates the evolution of five key indicators for the agricultural sector at the regency/city level. In general, the population has experienced a steady upward trend from around 1.55 million to nearly 1.8 million, reflecting continued population growth. Meanwhile, rice production declined sharply from 2017 to 2019 and has remained stagnant since then. The number of people working in agriculture decreased in 2015 and rose again in 2020, indicating a shift of labor to other sectors during that period. Harvested area shows a gradual downward trend, which may reflect a reduction in agricultural land or changes in cropping patterns. On the other hand, agricultural productivity has remained relatively stable without significant increases, which could indicate stagnation in technical innovation or external challenges faced by farmers. Overall, this graph suggests pressure on food security due to the increasing population, which is not accompanied by increases in production and harvested area.

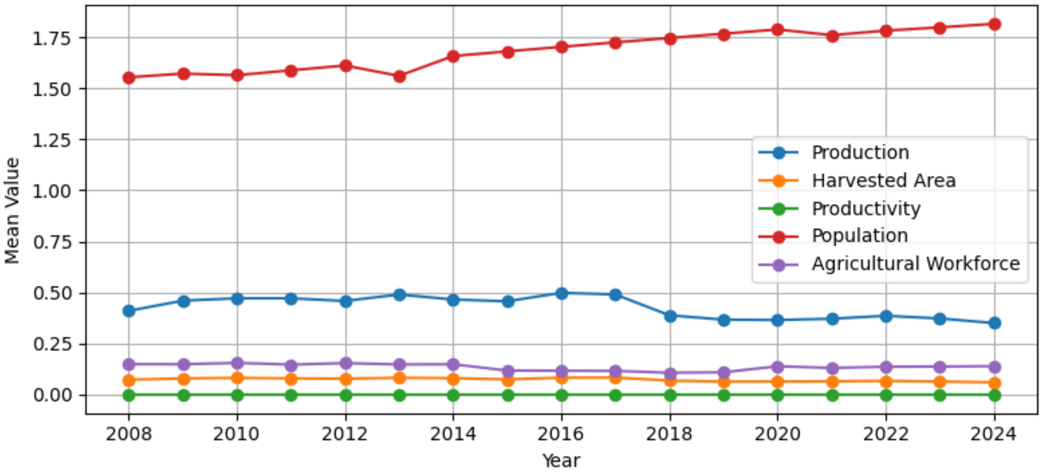


Figure 2. Trend of Mean Key Variables

Table 1 presents descriptive statistics that show significant disparities between regions, especially between cities and regencies. The average rice production per region reached approximately 427,835 tons, with a standard deviation of 402,050 tons, reflecting very high variations that ranged from 31 tons to more than 1.5 million tons. The harvest area also shows a similar pattern, with an average of 73,364 hectares and a range of 5 to 255,983 hectares. These differences reflect the dominance of regencies, which have larger areas and agricultural land compared to cities, which generally have limited agricultural land. Productivity is relatively more stable, with an average of 57.36 quintals per hectare and a median of 57.63, indicating a reasonably even distribution across regions, both cities and regencies. From a demographic perspective, the average population is approximately 1.68 million people, with a range of 17,000 to nearly 3.9 million people, reflecting the highly varied characteristics of the regions. The number of workers in the agricultural sector also varies significantly, ranging from 805 to 431,180 people, with an average of 135,241 people. This indicates that regencies are more dependent on the farming sector than cities. Overall, this data confirms that regencies tend to make a larger contribution to rice production, while cities play a minimal role due to land and labor constraints in agriculture.

Table 1. Descriptive Statistics of Primary Variables (2008–2024)

Variable	Mean	Std Dev	Min	25%	Median	75%	Max
Production (tons)	427,835.51	402,050.24	31.00	21,846.00	402,620.00	659,052.00	1,540,984.00
Harvest Area (ha)	73,364.23	67,566.77	5.00	3,527.00	71,635.00	113,697.00	255,983.00
Productivity (quintals/ha)	57.36	4.62	24.90	54.88	57.63	60.22	70.19
Population (people)	1,686,306.27	917,363.46	17,201.23	1,049,714.00	1,687,776.00	2,413,398.50	3,899,017.00
Agricultural Workforce (people)	135,241.78	117,321.01	805.00	10,603.00	128,332.00	235,232.50	431,180.00

The differences in values between regions, as shown in Table 1, reflect significant spatial variations in West Java, particularly between cities and regencies, in terms of rice production and supporting factors, including harvest area, productivity, population, and labor in the agricultural sector. These disparities are closely related to differences in geographical conditions, land capacity, and regional economic structures. Cities generally have limited farmland and make a low contribution to production, while regencies dominate agricultural activities. The wide distribution of data provides a crucial basis for applying clustering methods to group regions with similar characteristics.

4.2 Determining the Number of Clusters

Figure 3 shows that inertia decreases sharply from K = 2 to K = 5 and then flattens after K = 6, indicating an elbow point at K = 5. However, further evaluation metrics suggest that K = 6 provides a more optimal solution. The lowest DBI value (0.8221) was obtained at K = 6, while the Silhouette Score at K = 6 (0.4531) was slightly higher than at K = 5 (0.4462), indicating more coherent clustering.

Sensitivity tests using different initialization values (n_init = 10, 50, 100) and MinMaxScaler normalization consistently confirmed the superiority of K = 6 over K = 5, with lower inertia, higher Silhouette scores, and smaller DBI values (Table 2). These results demonstrate that although K = 5 represents the inertia elbow point, K = 6 offers the most stable, representative, and reliable clustering solution to capture the spatiotemporal dynamics of rice production in West Java.

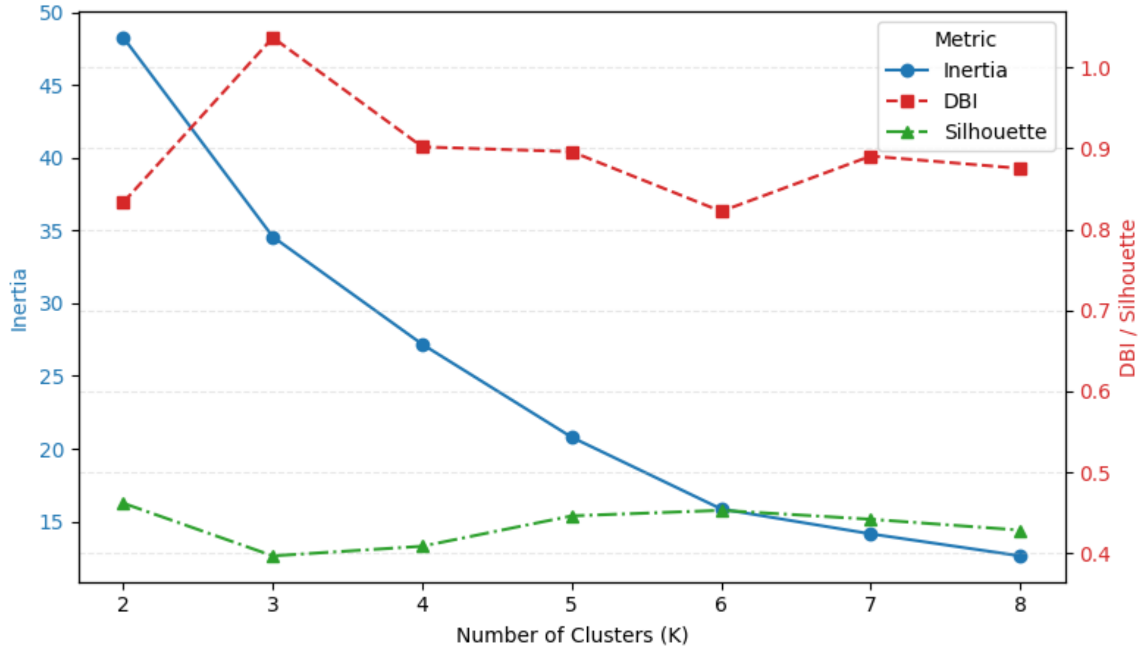


Figure 3. Cluster number evaluation

Table 2. Sensitivity analysis of clustering evaluation metrics for K=5 and K=6

n_init	K	Silhouette	DBI	Inertia
10	5	0.4462	0.8958	21
	6	0.4531	0.8221	16
50	5	0.4430	0.9132	21
	6	0.4531	0.8221	16
100	5	0.4430	0.9132	21
	6	0.4531	0.8221	16

4.3 PCA Cluster Visualization

Figure 4 is a visualization of clusters in a two-dimensional space resulting from PCA dimension reduction, showing a fairly clear separation between clusters. The six clusters formed (ranging from 0 to 5) are scattered across various positions on the PC1 and PC2 planes, which represent the two primary components of data variation. Clusters 1 and 5 are pretty separate from the other clusters, indicating distinctive and consistent characteristics. Clusters 3 and 4 also form dense and localized groups, indicating good internal compactness. Although there is some overlap between clusters 0, 2, and 3 in the central part of the graph, most data points within each cluster still maintain a sufficient distance from one another. This indicates that the clustering results not only successfully group regions with similar characteristics but also reflect the presence of spatiotemporal structure in the data.

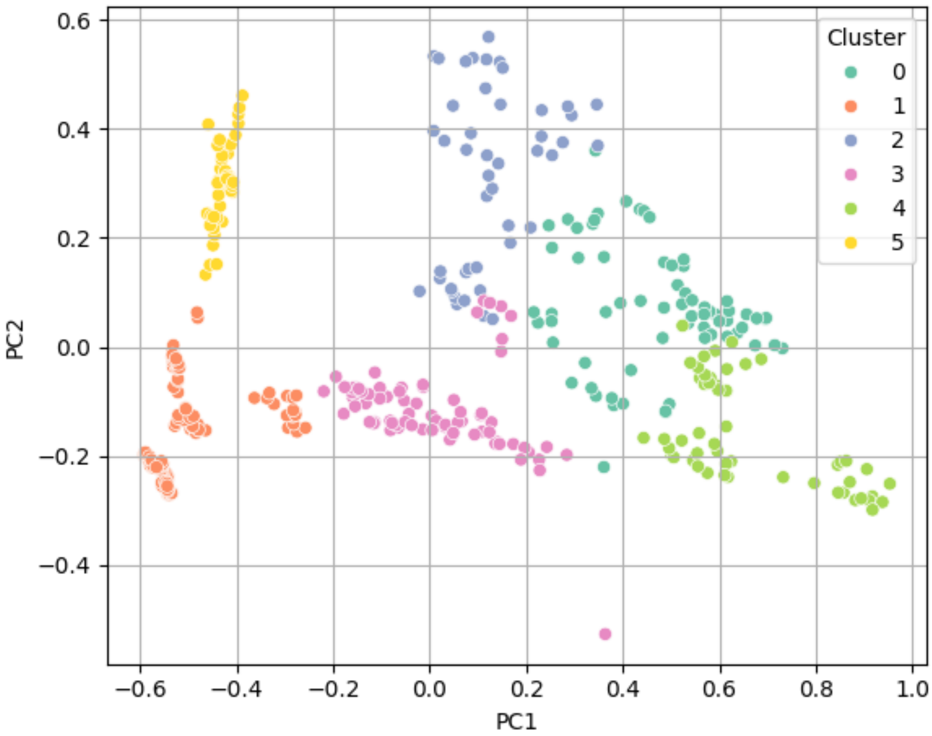


Figure 4. PCA Cluster Visualization

4.4 GIS-Based Cluster Visualization

Figure 5 illustrates the GIS-based clustering results for West Java regencies and cities, organized into six groups (Clusters 0–5). The spatial distribution is as follows:

a. Cluster 0: Sukabumi, Cianjur, Garut, and Tasikmalaya Regencies. These southern regions are characterized by strong agricultural areas with relatively large land resources.
b. Cluster 1: Banjar, Bogor, Cirebon, Sukabumi, and Tasikmalaya Cities, along with Purwakarta Regency. This cluster is dominated by autonomous cities and areas with relatively small contributions to rice production.
c. Cluster 2: Bandung, Bekasi, and Cirebon Regencies. These regions face high urbanization pressures, accompanied by a decline in agricultural engagement.
d. Cluster 3: Ciamis, Kuningan, Majalengka, and Sumedang Regencies. This group represents mid-level agricultural regions with balanced demographic and production characteristics.
e. Cluster 4: Indramayu, Karawang, and Subang Regencies. This cluster represents the main rice production centers in the northern belt.

f. Cluster 5: Bandung, Bekasi, and Depok Cities. This cluster consists of large urban cities with limited agricultural activity.

These spatial patterns reveal distinct differences between agricultural-dominated regions (southern and northern areas) and more urbanized areas (central and major cities). The clustering also reveals that autonomous cities are predominantly concentrated in Clusters 1 and 5, with limited agricultural contributions, while northern regencies dominate Cluster 4 as the primary rice production centers.

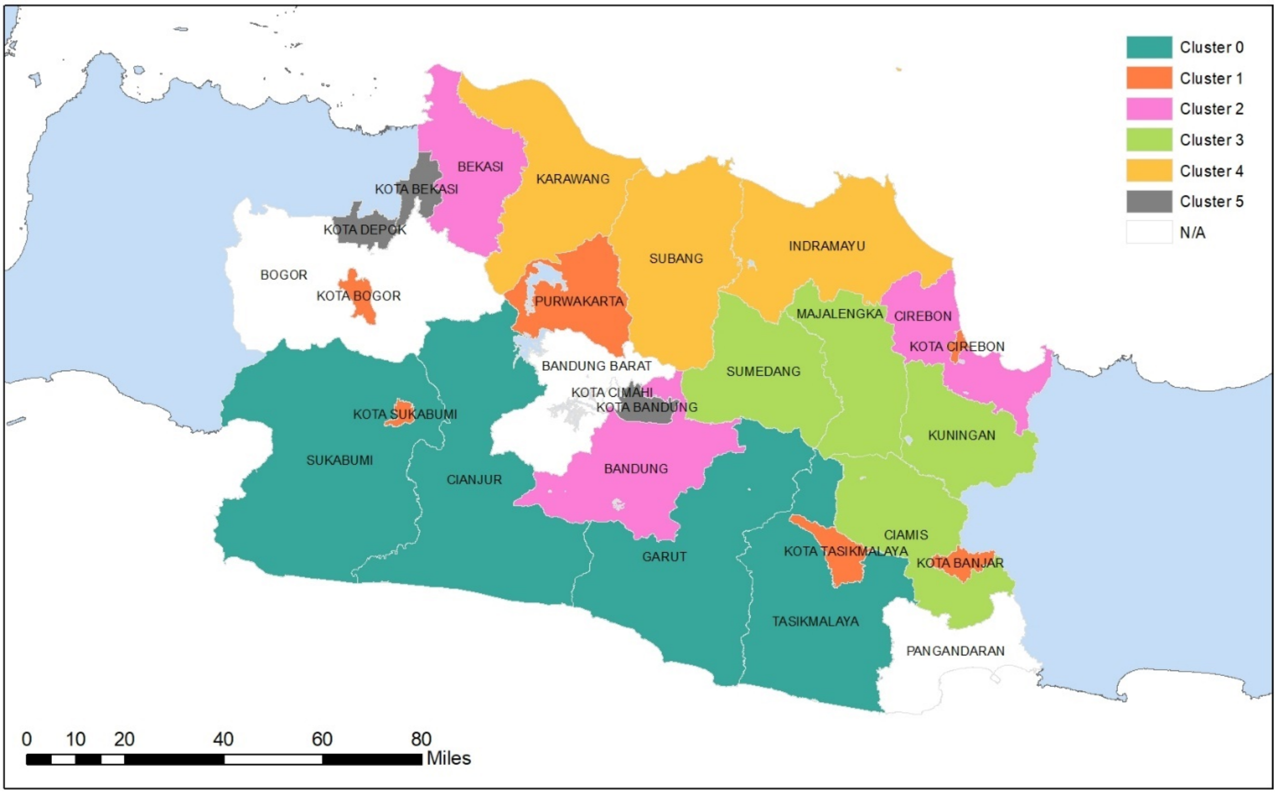


Figure 5. GIS Based Cluster Visualization

4.5 Characteristics of Each Cluster

The boxplot visualization in Figure 6 clearly shows differences in characteristics between clusters, indicating that the clustering successfully groups regions with distinct agricultural profiles. Cluster 4 consistently shows the highest values for production, harvest area, and productivity variables, indicating regions with large-scale farming operations. Cluster 5, on the other hand, has extremely low values for almost all variables, indicating regions with minimal agricultural activity. Clusters 1 and 3 tend to have relatively low production and harvest area values, but with different distributions in terms of population size and the number of people working in the agricultural sector. A high population characterizes cluster 2, but production and productivity are not particularly prominent, suggesting a potential disparity between population and production output.

Meanwhile, productivity shows a relatively even distribution across clusters, with values that are not significantly different. This suggests that differences in land productivity are not as significant as differences in total production volume, which are likely more influenced by harvest scale (area) and the number of agricultural workforce. Overall, this pattern suggests that clustering can effectively distinguish regions based on a combination of structural agricultural characteristics, rather than relying on a single variable.

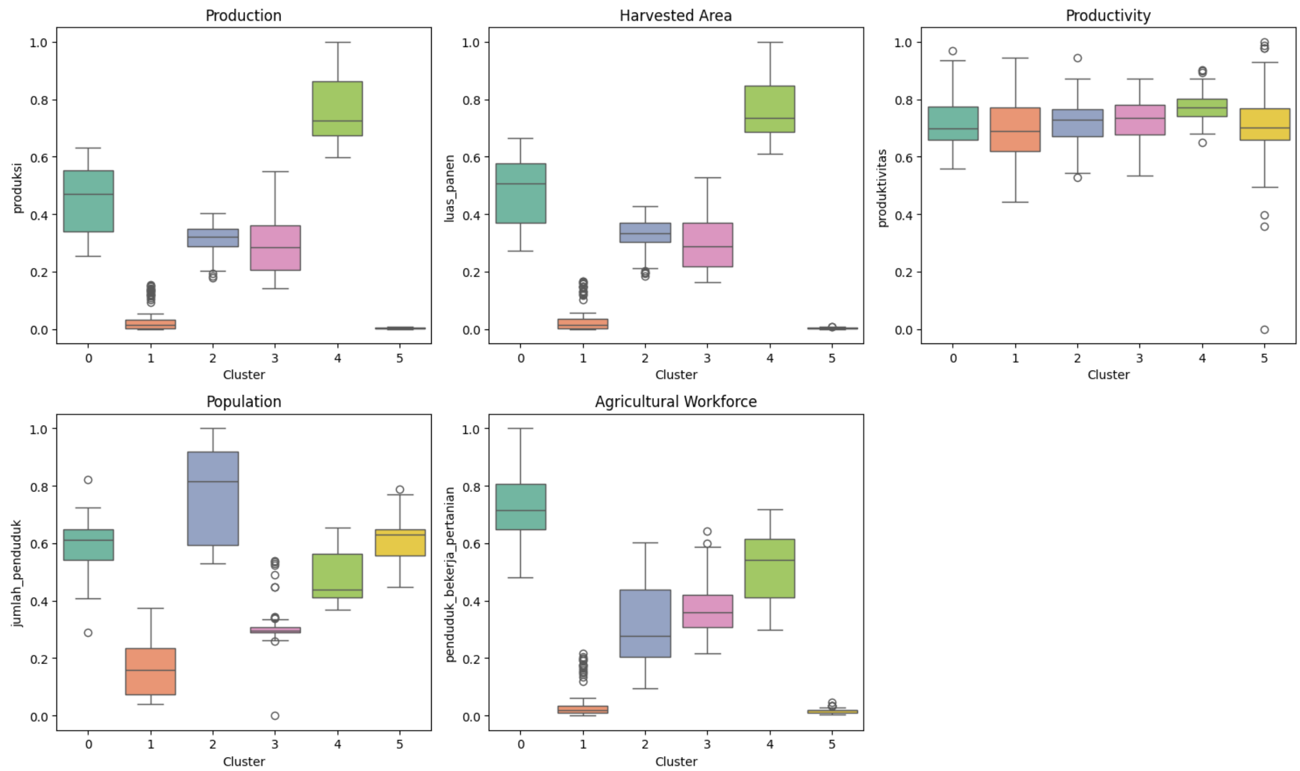


Figure 6. Variable Distribution in Each Cluster

4.6 Spatiotemporal Dynamics of Clusters

The spatiotemporal dynamics of rice production clusters in West Java were analyzed through multiple dimensions, including changes in cluster membership, transition patterns, regional stability, temporal heatmaps, and extreme changes. This analysis provides an overview of the consistency and significant changes in the characteristics of rice production in each region, offering insights into both regional stability and vulnerability.

4.6.1 Number of Regencies per Cluster from Year to Year

Figure 7 shows the dynamics of the number of regencies/cities in each cluster during the period 2008 to 2024. The fluctuation pattern reflects shifts in regional characteristics over time based on the agricultural indicators analyzed.

Cluster 1 dominated throughout the period, with a stable number of regencies in the range of 6 to 7 regions. This suggests that most areas exhibit similar and relatively stable characteristics, aligning with the profile of cluster 1. Clusters 0 and 3 show greater fluctuations. Cluster 0 experienced a significant decline after 2010, while Cluster 3 increased rapidly until 2013 before finally declining and stabilizing. Clusters 2 and 4 tend to have fewer and more fluctuating regencies, indicating characteristics that are less common but still relevant during specific periods. Cluster 5 shows high consistency, maintaining the number of regencies at three since 2010, indicating the presence of a group of regions with a particular and essentially unchanged profile.

The stability of the number of regencies in some clusters (such as Clusters 1 and 5) reflects consistent spatiotemporal characteristics. In contrast, fluctuations in other clusters reflect agricultural dynamics, such as changes in policy, agricultural technology, or socio-economic conditions that influence production, harvested area, and the number of people in the agricultural workforce.

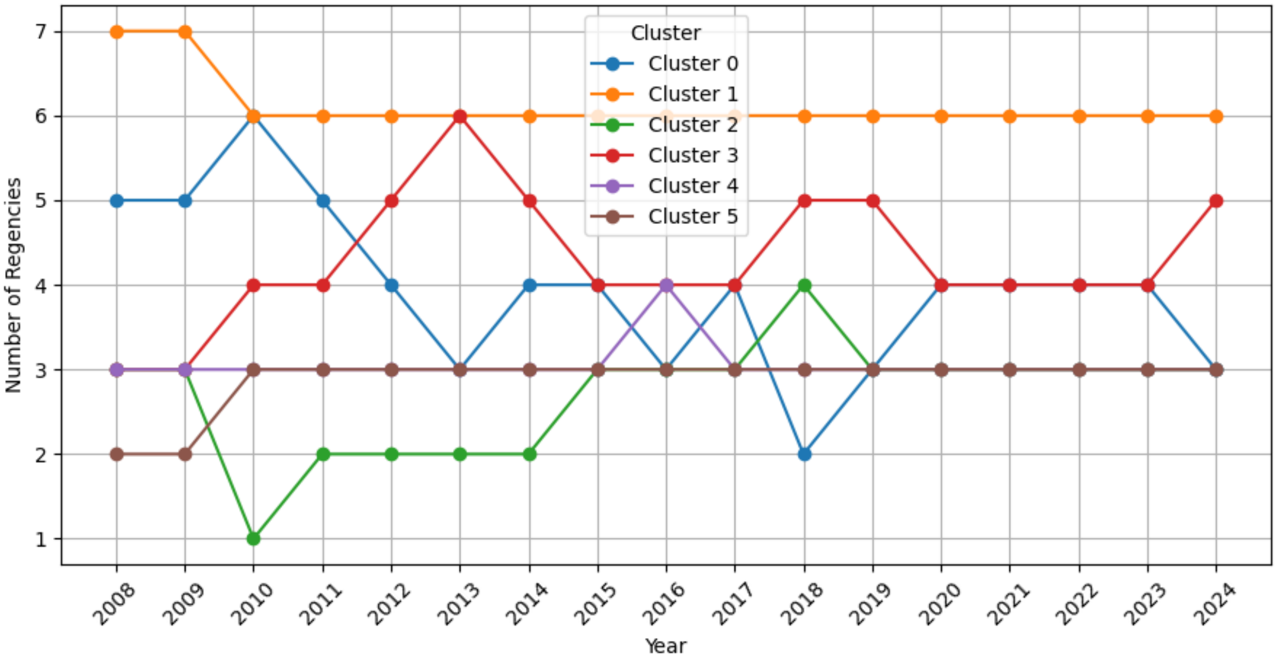
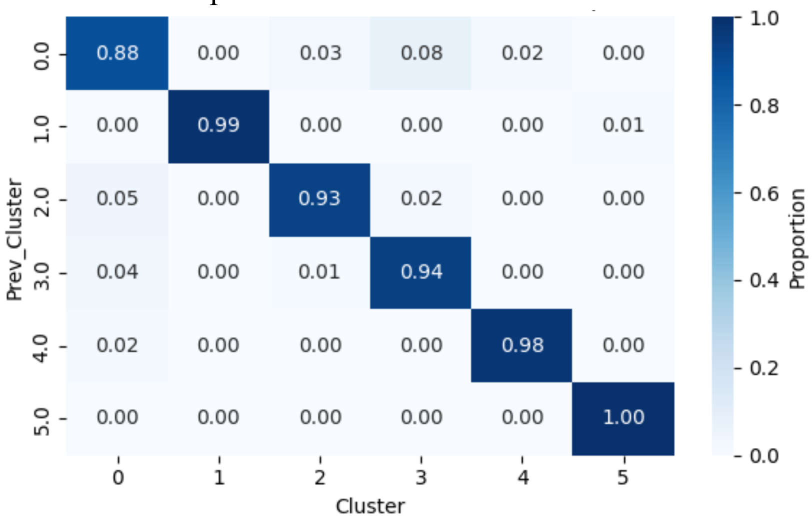


Figure 7. Number of Regencies per Cluster (2008–2024)

4.6.2 Transition Patterns Between Clusters

Figure 8 is a year-on-year cluster transition heatmap illustrating the proportion of regencies/cities that transitioned or remained in the same cluster from one year to the next. High central diagonal values indicate strong cluster stability, with Cluster 5 exhibiting perfect stability (1.00), followed by Cluster 1 (0.99), Cluster 4 (0.98), Cluster 3 (0.94), Cluster 2 (0.93), and Cluster 0 (0.88). This indicates that most regions have relatively consistent agricultural characteristics over time. Meanwhile, the small off-diagonal proportions, such as the 8% transition from Cluster 0 to Cluster 3 and the 5% transition from Cluster 2 to Cluster 0, indicate spatiotemporal dynamics in certain regions. Although most areas show stability, these inter-cluster movements suggest structural changes or external influences affecting the agricultural sector in specific areas.



Prev_Cluster	0	1	2	3	4	5
0.0	0.88	0.00	0.03	0.08	0.02	0.00
1.0	0.00	0.99	0.00	0.00	0.00	0.01
2.0	0.05	0.00	0.93	0.02	0.00	0.00
3.0	0.04	0.00	0.01	0.94	0.00	0.00
4.0	0.02	0.00	0.00	0.00	0.98	0.00
5.0	0.00	0.00	0.00	0.00	0.00	1.00

Figure 8. Year-to-Year Cluster Transition Heatmap

4.6.3 Regencies Cluster Stability

The histogram in Figure 9 shows that most regencies have a very high level of stability, with 16 regencies (the majority) falling within the 100% stability range, meaning they never changed clusters during the observation period. Several other regencies showed stability between 90% and 95%, and only a small number fell below 90%, with some even having stability of only around 70–75%. This distribution indicates that the majority of regions maintain consistent agricultural characteristics from year to year. At the same time, a small portion of areas experience more dynamic changes, which may be attributed to variations in production factors, policy changes, or local socio-economic conditions.

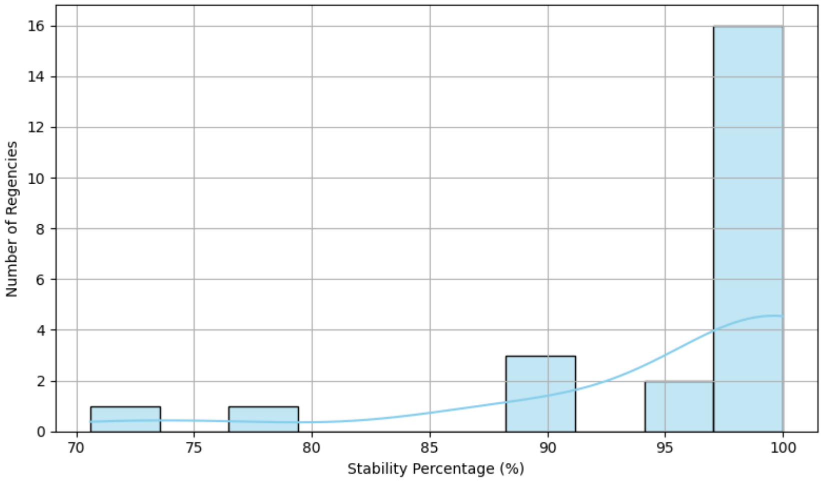


Figure 9. Histogram of Cluster Stability Distribution by Regency

4.6.4 Heatmap of Regencies/Cites Clusters by Year

Figure 10 shows the dynamics of cluster membership, with each row representing one regency/city and each column representing the year of observation. Different colors represent six clusters (from 0 to 5), and color changes indicate the movement of regencies/cities from one cluster to another.

Most regencies/cities, such as Sukabumi Regency, Kuningan Regency, Indramayu Regency, Bekasi Regency, Bogor City, and Bandung City, exhibit consistent colors over time, indicating that their membership in a single cluster remains relatively stable. Conversely, some regencies, such as Tasikmalaya and Cirebon, exhibit more varied color patterns, reflecting more frequent transitions between clusters, which significant changes in production factors, harvest areas, or local socio-economic conditions may cause. This pattern reinforces previous findings that the majority of areas in West Java exhibit high spatiotemporal stability, while others exhibit dynamics that warrant further attention.

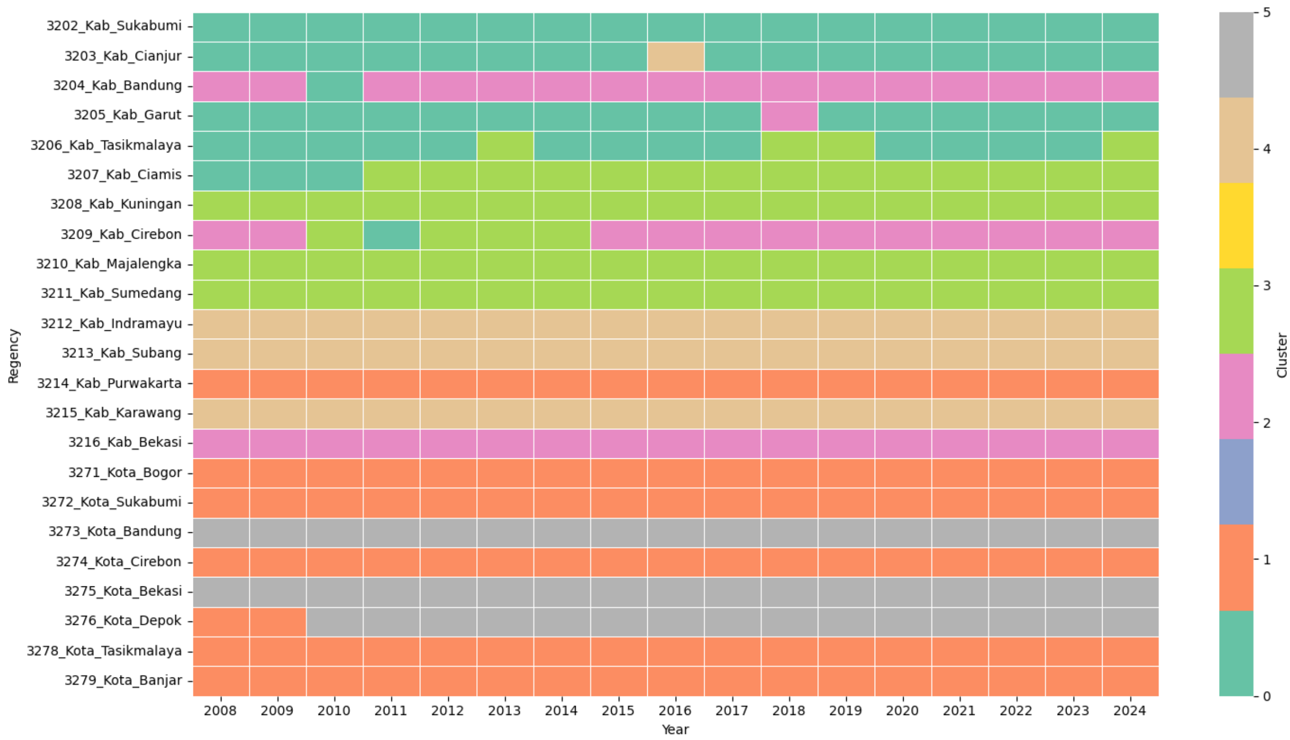


Figure 10. Heatmap of Regencies/Cities Clusters by Year

4.6.5 Extreme Cluster Changes: Case Studies of Bandung and Garut Regencies

Bandung and Garut Regencies showed striking spatiotemporal dynamics in terms of changes in extreme clusters between clusters 0 and 2 during the observation period. Bandung Regency moved from cluster 2 to cluster 0 in 2010, then returned to cluster 2 in 2011. Meanwhile, Garut Regency shifted from cluster 0 to 2 in 2018 and returned to cluster 0 in 2019. These significant changes indicate instability in regional characteristics, which may be caused by external factors such as climate fluctuations, agricultural policies, or disruptions to infrastructure. Both regencies warrant further analysis to understand the causes of these fluctuations and to formulate data-driven strategies for managing agricultural risks.

Overall, the findings suggest that while most regencies in West Java exhibit relatively stable agricultural characteristics over time, specific regions, such as Bandung and Garut Regencies, experience substantial fluctuations that may signal structural vulnerabilities. This highlights the need to translate clustering outcomes into concrete policy recommendations, ensuring that interventions are both targeted and adaptive to regional conditions.

4.7 Policy Implications of Clustering Results

Building upon the clustering analysis and spatiotemporal dynamics presented in the previous sections, this subsection translates the empirical findings into actionable policy directions. Each cluster represents distinct agricultural characteristics that necessitate tailored strategies to enhance productivity, ensure regional stability, and strengthen resilience. Table 3 summarizes suggested intervention strategies for each cluster type.

Table 3. Policy Recommendations Based on Cluster Characteristics

Cluster	Regencies/Cities	Characteristics	Policy Recommendations
0	Sukabumi, Cianjur, Garut, Tasikmalaya	Vulnerable to production fluctuations, relatively lower stability	Climate risk mitigation, adaptive subsidies, farmer capacity building, regular monitoring
1	Banjar City, Bogor City, Cirebon City, Sukabumi City, Tasikmalaya City, Purwakarta	Small-scale, semi-urban, low agricultural land	Support for urban farming, optimize urban–rural linkages, promote alternative livelihoods
2	Bandung, Bekasi, Cirebon	Medium-scale, relatively stable, moderate contribution	Balanced interventions: sustainable intensification and diversification programs
3	Ciamis, Kuningan, Majalengka, Sumedang	Moderate productivity, unstable dynamics	Diversification of commodities, improved irrigation management, strengthened farmer cooperatives
4	Indramayu, Karawang, Subang	Stable high-production centers	Focus on intensification, modernization (mechanization, precision agriculture), supply chain strengthening
5	Bandung City, Bekasi City, Depok City	Highly urbanized, minimal direct agricultural activity	Urban food security programs, agri-tech innovation (vertical farming, hydroponics), urban–rural integration policies

4.8 Discussion

4.8.1 Regional Segmentation and Cluster Representation

The clustering results successfully mapped significant spatial variations among regencies/cities in West Java. Cluster 4, for example, includes areas such as Indramayu and Karawang, which consistently rank high in terms of production, harvest area, and population working in the agricultural sector. This reflects the characteristics of national rice production centers, which are supported by technical irrigation infrastructure and efficient distribution. Conversely, Cluster 5 encompasses regions with significant limitations, both in terms of production and harvested area, including the cities of Bandung and Bekasi. The relatively uniform distribution of productivity variables across clusters indicates that land efficiency is not the primary differentiator, and production variations are more influenced by harvest scale and labor capacity. The resulting segmentation not only identifies absolute patterns in production figures but also highlights the balance between population, labor input, and agricultural production output.

4.8.2 Spatiotemporal Dynamics and Regional Instability

Temporal analysis reveals that most regions exhibit high stability, as indicated by the transition heatmap and stability histogram. Cluster 5 maintained perfect stability during 17 years of observation, and cluster 1 approached a similar value (0.99). This indicates that many regions in West Java have consistent agricultural profiles over time. However, some regions exhibit extreme dynamics, such as

Bandung and Garut Regencies, which have moved between clusters 0 and 2 in a short period. This phenomenon suggests potential sensitivity to external factors, such as land conversion in urban areas (Bandung Regency) or variability in planting seasons and climate (Garut Regency).

4.8.3 Policy Implications and Recommendations

Spatiotemporal segmentation enables a more contextual approach to intervention. Areas with high stability can be focused on for location-specific technology development and production intensification. Conversely, dynamic regions require adaptive strategies and regular monitoring, such as planning for fertilizer distribution and mitigating climate change risks. Policy recommendations can also be directed toward developing cluster-based budgets that take regional dynamics into account, thereby improving the efficiency and accuracy of interventions.

4.8.4 Limitations and Further Study

Although interpolation was used to address data gaps (e.g., 2016 for all regencies and 2011 for Cirebon City), the estimates remain uncertain. This method assumes a linear trend and may introduce potential bias. Alternative approaches, such as multiple imputation or time-series-specific methods (e.g., ARIMA-based or seasonal decomposition), could provide more robust imputations. These methods were not applied in this study due to computational and data limitations, but they are acknowledged as possible directions for future work.

In addition, the use of K-Means is limited in recognizing non-linear cluster shapes and complex distributions; therefore, further studies are recommended to adopt alternative methods, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) or Hierarchical DBSCAN, which are more robust to uneven cluster shapes and densities. The development of time-series clustering models or the integration of spatial environmental data, such as climate and topography, will also enrich the approach to agricultural area segmentation in the future.

5 Conclusion

This research classified 23 regencies and cities in West Java (2008–2024) using K-Means clustering based on five variables: production, harvested area, productivity, population, and agricultural workforce. Evaluation metrics (Elbow, Silhouette, and DBI) confirmed six optimal clusters, which PCA and GIS-based visualization supported.

Results show precise regional segmentation: Indramayu, Karawang, and Subang regencies remain stable high-production centers; Bandung and Garut regencies display dynamic shifts driven by urbanization and climate variability; while urban cities such as Bandung, Bekasi, and Depok contribute minimally, requiring urban–rural food security strategies. Most regions demonstrated long-term stability, making them suitable for intensification and modernization, while more dynamic areas require adaptive risk-mitigation measures. Urban regions should focus on agricultural innovation.

Despite limitations in data interpolation and K-Means' inability to capture non-linear clusters, this research highlights the value of integrating spatiotemporal clustering and GIS for policy-relevant insights in agriculture. Future research should incorporate climate, topography, and land-use data with more advanced clustering methods. These include algorithms like DBSCAN and its extension, HDBSCAN, the latter being superior for modeling clusters with varying densities and capturing irregular, non-linear relationships.

Overall, the findings provide an evidence-based framework to support adaptive and targeted agricultural policies, strengthening regional and national food security.

Acknowledgement

This manuscript received language assistance from OpenAI ChatGPT-4o, with all outputs reviewed and revised by the author to maintain academic integrity and clarity.

Bibliography

[1] A. Musyafak, A. Susanti, A. Supriyatna, T. H. A. Takariyana, V. S. Bonavia, and Suyati, Outlook Komoditas Pertanian Tanaman Pangan: Padi. Jakarta: Pusat Data dan Sistem Informasi Pertanian, Sekretariat Jenderal Kementerian Pertanian, 2020.

[2] E. T. Tosida, F. D. Wihartiko, I. Hermadi, Y. Nurhadryani, and Feriadi, "The Big Data Commodity Management Model for Rice for National Food Policy," J. RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 4, no. 1, pp. 142–154, Feb. 2020, doi: 10.29207/resti.v4i1.1520.

[3] Antara News, "Pemkab Indramayu catat provitas padi capai 7,3 ton per hektare," Antaranews.com, 5 Jul. 2024. [Online]. Available: https://www.antaranews.com/berita/4089546/pemkab-indramayu-catat-provitas-padi-capai-73-ton-per-hektare

[4] Badan Pusat Statistik Provinsi Jawa Barat, "Luas Panen, Produktivitas, dan Produksi Padi Menurut Kabupaten/Kota di Provinsi Jawa Barat," BPS Jawa Barat, 2024. [Online]. Available: https://jabar.bps.go.id/id/statistics-table/

[5] M. Murdiana and F. Fadli, "Peran irigasi dalam peningkatan produksi padi sawah di Kecamatan Meurah Mulia Kabupaten Aceh Utara," Agrifo: Jurnal Agribisnis Universitas Malikussaleh, vol. 1, no. 2, 2016, doi: 10.29103/ag.v1i2.760.

[6] Septi, "Faktor-faktor yang memengaruhi produksi padi di Kabupaten Tapanuli Selatan," PROFJES: Profetik Jurnal Ekonomi Syariah, vol. 2, no. 2, Jul.–Des. 2023. [Online]. Available: http://jurnal.iain-padangsidimpuan.ac.id/index.php/Profetik

[7] M. R. I. Amar, Faktor-faktor yang mempengaruhi hasil produksi padi di Kabupaten Pinrang, Skripsi, Univ. Islam Negeri Alauddin Makassar, 2023.

[8] R. P. Hidayat, E. T. Tosida, A. Maesya, and I. P. Solihin, "Support vector regression algorithm for predicting rice production in West Java Province," Proc. Int. Conf. on Informatics, Multimedia, Cyber and Information System (ICIMCIS), IEEE, 2024, doi: 10.1109/ICIMCIS57984.2024.10390639.

[9] J. Hutahaean, D. Yusup, and Purwantoro, "Perbandingan metode Linear Regression, Random Forest & K-Nearest Neighbor untuk prediksi produksi hasil panen padi di Provinsi Jawa Barat," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 3, Jun. 2024, doi: 10.36040/jati.v8i3.9821

[10] S. K. Dermoredjo et al., "National food development policies in Indonesia: An analysis of food sustainability and security," BIO Web of Conferences, vol. 119, 05006, 2024, doi: 10.1051/bioconf/202411905006.

[11] P. Kurniawati, H. Pratiwi, and Sugiyanto, "Indonesian territory clustering based on harvested area and rice productivity using clustering algorithm," J. Social Sci., vol. 4, no. 1, 2023, doi: 10.46799/jss.v4i1.510.

[12] Abirami et al., "Application of K-means Clustering Algorithm in Rice Production of Tamil Nadu, India," Int. J. Environ. Climate Change, vol. 12, no. 11, pp. 1348–1355, 2022, doi: 10.9734/IJECC/2022/v12i1131113.

[13] L. Wickramasinghe et al., "Modeling the relationship between rice yield and climate variables using statistical and machine learning techniques," J. Mathematics, vol. 2021, Art. ID 6646126, 2021, doi: 10.1155/2021/6646126.

[14] R. Savira, "Optimizing clustering models using principle component analysis for car customers," IJCCS (Indonesian J. Comput. Cybern. Syst.), vol. 18, no. 2, Apr. 2024, doi: 10.22146/ijccs.94744.

[15] N. Omar et al., "Integration of principal component analysis and K-means clustering for type 2 diabetes sub-clustering model," AIP Conf. Proc., vol. 3056, no. 1, p. 070001, Apr. 2025, doi: 10.1063/5.0209965.

[16] E. Umargono, J. E. Suseno, and S. K. V. Gunawan, "K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula," Proc. 2nd Int. Seminar on Science and Technology (ISSTEC), Oct. 2020, doi: 10.2991/assehr.k.201010.019.

[17] F. Tempola and A. F. Assagaf, "Clustering of Potency of Shrimp in Indonesia With K-Means Algorithm and Validation of Davies-Bouldin Index," Proc. Int. Conf. on Science and Technology (ICST), Dec. 2018, doi: 10.2991/icst-18.2018.148.

[18] M. Joseph et al., "Modelling climate variabilities and global rice production: A panel regression and time series analysis," Heliyon, vol. 9, no. 11, e15480, 2023, doi: 10.1016/j.heliyon.2023.e15480.

[19] S. M. Algarni et al., "Analyzing the impact of climate change on rice production and strategies for enhancing efficiency, sustainability, and global food security," Int. J. Innov. Res. Sci. Stud., vol. 8, no. 2, pp. 2946–2957, 2025, doi: 10.53894/ijirss.v8i2.5888.

[20] K. Suresh et al., "How productive are rice farmers in Sri Lanka? The impact of resource accessibility, seed sources and varietal diversification," Food Security, vol. 15, pp. 1005–1024, 2023, doi: 10.1007/s12571-023-01338-2.

[21] H. Zhu et al., "Spatio-temporal multi-level attention crop mapping method using time-series SAR imagery," ISPRS J. Photogramm. Remote Sens., 2023, doi: 10.1016/j.isprsjprs.2023.11.016.

[22] Supriatna et al., "Spatio-temporal analysis of rice field phenology using Sentinel-1 image in Karawang Regency West Java, Indonesia," Int. J. GEOMATE, vol. 17, no. 62, pp. 101–106, Oct. 2019, doi: 10.21660/2019.62.8782.

[23] R. S. V. Teegavarapu, "Temporal Interpolation Methods," in Imputation Methods for Missing Hydrometeorological Data Estimation, Water Sci. Technol. Libr., vol. 108, Springer, Cham, 2024, doi: 10.1007/978-3-031-60946-6_3.

[24] Wongoutong, "The impact of neglecting feature scaling in k-means clustering," PLOS ONE, vol. 19, no. 12, e0310839, Dec. 2024, doi: 10.1371/journal.pone.0310839.

[25] M. Gaido, "Distributed Silhouette: Efficient Computation of the Silhouette Score on Large Datasets," arXiv preprint, arXiv:2303.14102, 2023.

[26] B. Sadeghi, "Clustering in geo-data science: Navigating uncertainty to select the most reliable method," Ore Geol. Rev., vol. 181, 106591, 2025, doi: 10.1016/j.oregeorev.2025.106591.

[27] J. A. Hartigan and M. A. Wong, "A K-Means clustering algorithm," J. Royal Stat. Soc. Series C (Appl. Stat.), vol. 28, no. 1, pp. 100–108, Mar. 1979, doi: 10.2307/2346830.