Available online at https://ejournal.corespub.com/index.php/ijmsc/index

International Journal of Mathematics, Statistics, and Computing

e-ISSN 3025-0803

Vol. 3, No. 4, pp. 159-164, 2025

Clustering of Regencies and Cities in West Java Province Based on Horticultural Indicators Using the K-Means Method

Rivani Lakui[1*], Gerry Sastro[2], Tabah Heri Setiawan[3]

[1,2,3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Pamulang University, Tangerang Selatan, Banten, Indonesia*

**Corresponding author email: rivani.lakui@student.universitaspamulang.ac.id*

Abstract

National food security largely depends on the capacity of domestic production. However, Indonesia continues to rely on food imports, including horticultural products. West Java Province, as one of the country's main food-producing regions, possesses diverse geographical conditions that support the cultivation of various commodities and therefore becomes the focus of this study. This research aims to classify the 27 districts and cities of West Java Province based on horticultural indicators in order to identify spatial patterns and development potential. The study employed secondary data from the Central Bureau of Statistics (BPS) of West Java and World Climate for 2023, including horticultural production (ornamental plants, bio-pharmaca, vegetables, and fruits), annual average rainfall, and temperature. The analysis used the K-Means clustering method, with the Silhouette Index as an evaluation measure to determine the optimal number of clusters. Results indicate that three clusters provided the best accuracy. Cluster 1 (e.g., Cianjur and Bandung Regencies) consists of areas with high horticultural production, low temperatures, and moderate rainfall. Cluster 2 (e.g., Bogor and Sukabumi Regencies) represents regions with low production, moderate temperature, and high rainfall. Cluster 3 (e.g., Cirebon and Indramayu Regencies) includes areas with moderate production, high temperature, and low rainfall. The findings provide a foundation for local and national governments to design targeted horticultural development strategies that enhance productivity, improve farmers' welfare, and support sustainable food security.

Keywords: Clustering, K-Means, Silhouette Index, Food Security, Horticulture

1. Introduction

Food security is a fundamental necessity for human survival, and in Indonesia, it is strongly tied to the performance of the agricultural sector (Timmer, 2004). According to Singh et al. (2024), agriculture not only ensures food supply but also drives economic growth, alleviates poverty, and sustains rural livelihoods. Consequently, the sustainability of agriculture is crucial for economic and social stability. Despite its importance, Indonesia's self-sufficiency in food has been inconsistent (Simatupang & Peter Timmer, 2008). After achieving rice self-sufficiency in 1984, the country resumed rice imports in 1988, influenced by land conversion, climate change, declining interest in farming among younger generations, and limited adoption of agricultural technology. The urgency of this issue is starkly highlighted by recent data; in 2024, food imports reached 4.52 million tons, the highest in seven years, underscoring the critical need to strengthen national food production.

Within the agricultural sector, horticulture plays a strategic role (Saha et al., 2021). It contributed 1.44% to national GDP in 2023, provides daily nutrition, has high economic value, and carries significant export potential. Indonesia is the largest horticultural producer in ASEAN, with production exceeding 34 million tons, surpassing Vietnam, Philippines, and Thailand. However, a deeper analysis reveals a significant challenge; when adjusted for land area, Indonesia's productivity lags behind these smaller countries, reflecting inefficiencies in land management and technological utilization (Ruslan, 2021).

West Java Province is one of Indonesia's primary food producers, supported by diverse climatic and geographical conditions. The region's agricultural efficiency is enhanced through the use of combined synthetic and organic fertilisers, which improve paddy productivity compared to conventional farming (Hendrani et al., 2022). Nonetheless, horticultural productivity in the region remains suboptimal and is still dominated by vegetables, while other high-value commodities remain underdeveloped. However, a critical gap exists in current research. There has been little effort to systematically classify districts and cities in West Java based on key horticultural indicators. This spatial analysis is essential to move beyond a one-size-fits-all policy approach and is vital for capturing distinct regional patterns and potentials, which is the foundational step toward more effective and targeted policy design.

Therefore, to address this gap, this study aims to cluster the districts and cities of West Java Province using the K-Means method. The analysis will be based on three pivotal variables: horticultural production, rainfall, and average temperature. These variables were selected because production is a direct measure of output, while rainfall and temperature are fundamental climatic factors that dictate crop suitability, growth cycles, and ultimately, productivity. The resulting classification is expected to provide objective insights into regional similarities and differences, serving as a foundational tool for strategic horticultural development and sustainable food security in the region.

2. Literature Review

2.1. Previous Studies

With the rapid development of technology and the demand for accurate data analysis, the K-Means method has been widely applied across various fields of research. Several previous studies provide insights into how this method has been utilized to classify data based on similarities among characteristics.

For instance, Tadi & Ningsi (2023) examined poverty issues in Banten Province by clustering districts and cities based on socio-economic variables such as education, employment sectors, and food expenditure. They employed K-Means clustering combined with Silhouette Index evaluation to validate the results. Their findings revealed two distinct poverty clusters and highlighted the importance of addressing multicollinearity prior to analysis to ensure data quality.

Similarly, Lianita et al. (2023) applied K-Means clustering in the agricultural sector to map vegetable productivity in North Sumatra. By integrating K-Means with a web-based geographic information system, they classified crops such as potatoes, cabbage, and spinach into low, medium, and high productivity clusters. The visualization provided practical value by enabling local governments to easily access and utilize the results for decision-making.

Christiani (2024) focused on rice production in East Nusa Tenggara (NTT). Using survey data on harvested area, production volume, and productivity, the province was clustered into three categories: low, medium, and high potential. The study further applied ANOVA testing to statistically validate the significant differences among clusters, thereby strengthening the robustness of the results.

2.2. Comparison with This Study

The three aforementioned studies demonstrate the versatility of K-Means across different contexts. While Tadi & Ningsi (2023) focused on socio-economic factors, Lianita et al. (2023) concentrated on horticultural crops, and Christiani (2024) analyzed rice production. Despite similarities in methodology, their approaches varied in terms of data preparation, evaluation metrics, and additional statistical tests. Compared to these studies, the present research emphasizes horticultural indicators in West Java Province, with a specific focus on combining K-Means clustering with Silhouette Index evaluation. This approach is designed to ensure both computational efficiency and cluster validity, thereby offering more reliable insights for regional development planning.

2.3. Research Gap

Although K-Means clustering has been extensively used in socio-economic and agricultural contexts, several gaps remain. First, not all studies integrate cluster validation measures such as the Silhouette Index to ensure the quality of clusters. Second, applications in horticulture particularly in relation to food security and production planning remain relatively limited. Third, practical implications of clustering results are often underexplored, reducing their usefulness for policymakers and practitioners. This study addresses these gaps by applying K-Means clustering to horticultural data from West Java Province, standardizing variables of different scales, and validating the results using the Silhouette Index. The objective is not only to identify regional patterns but also to provide actionable insights for sustainable horticultural development.

3. Materials and Methods

3.1. Materials

The data used in this study were secondary sources obtained from the Central Bureau of Statistics (BPS) of West Java and World Climate for 2023. Variables include horticultural production (vegetables, fruits, ornamental plants, and bio-pharmaca), average annual rainfall, and average annual temperature for all 27 districts and cities in West Java

Province. Due to differences in measurement units (e.g., quintals, trees, millimeters, and °C), all variables were standardized prior to analysis to ensure comparability. The study area covers 19 districts and 9 cities across the province. The tools used in the analysis were RStudio and Microsoft Excel 2016.

3.2. Methods

The study employed cluster analysis using the K-Means method. The steps followed were as follows:

a. Data collection from BPS and World Climate.
b. Data cleaning and verification.
c. Standardization of variables with different units.
d. Selection of k (number of clusters) to be formed.
e. Initialization of centroids.
f. Calculation of distances between objects and centroids using Euclidean distance.
g. Recalculation of new centroids until convergence.
h. Validation of results using the Silhouette Index.

The research flow diagram is illustrated in Figure 1.

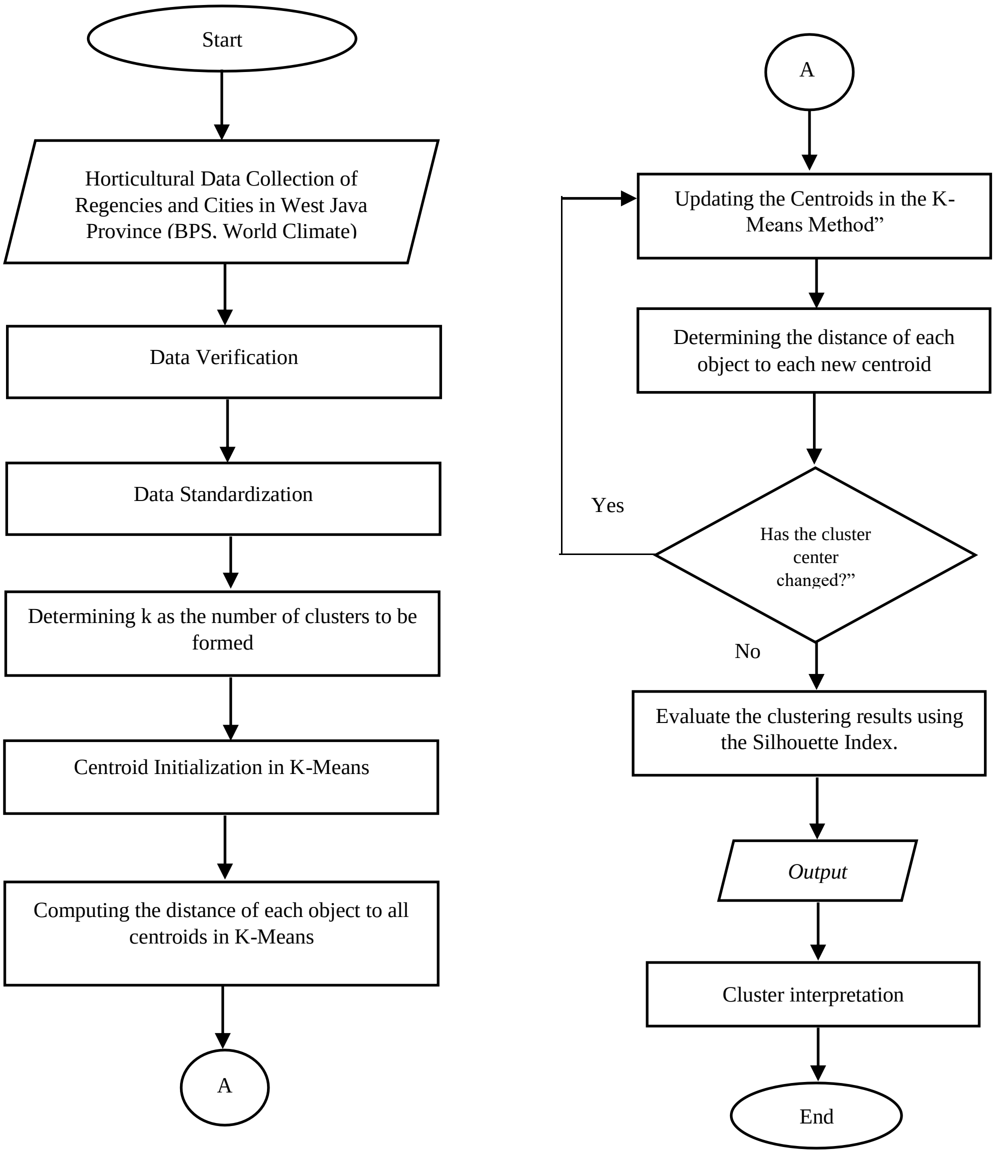

Figure 1: Research Flow Diagram

3.3. Data Standardization

Data standardization was applied to normalize variables with different units into z-scores by subtracting the mean and dividing by the standard deviation. This ensured that all variables contributed equally to the clustering process. The formula for data standardization is given as follows (Faisal & Rifai, 2023):

$$Z = \frac{x_i - \overline{x}}{s}, \tag{1}$$

where,

Z : Standardized value
x_i : Value of the i-th data point
$\bar{x}$: Mean of the dataset
s : Standard deviation

3.4. Silhouette Index

The Silhouette Index was used to validate cluster quality by measuring cohesion and separation among data points. A value close to 1 indicates strong structure, while values closer to 0 suggest weaker clustering. According to Kaufman & Rousseeuw (1990), Silhouette Index values can be interpreted as follows: $0.71 - 1.00$ (strong), $0.51 - 0.70$ (good), $0.25 - 0.50$ (weak), ≤ 0.25 (not structured).

Table 1: Silhouette Index Criteria

Silhouette Index Value	Interpretation
0.71-1.00	Strong structure
0.51-0.70	Good structure
0.25-0.50	Weak structure
≤ 0.25	Not structured

4. Results and Discussion

The analysis compared clustering solutions with $k = 2$, $k = 3$, and $k = 4$. Based on Silhouette Index values, the three-cluster solution achieved the highest validity score (0.37), compared to two clusters (0.33) and four clusters (0.36). This indicates that three clusters best represent the structure of horticultural data in West Java Province.

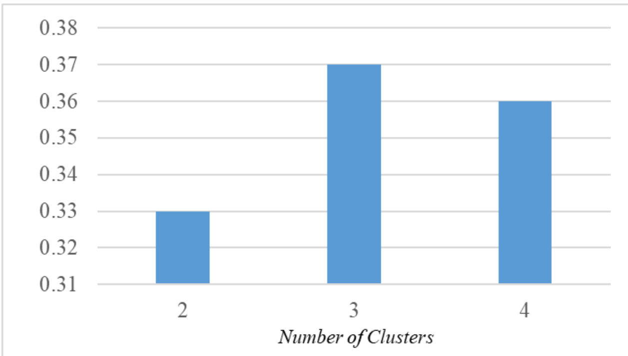

Figure 2: Silhouette Index Evaluation Results

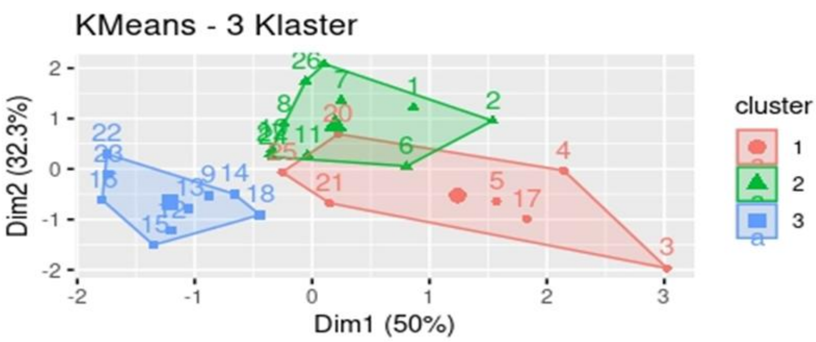

Figure 3: Cluster Plot with Three Clusters

Cluster 1 includes Cianjur Regency, Bandung Regency, Garut Regency, Bandung Barat Regency, Sukabumi City, and Bandung City, characterized by high horticultural production, low temperatures, and moderate rainfall, generally corresponding to highland areas. Cluster 2 consists of Bogor Regency, Sukabumi Regency, Tasikmalaya Regency, Ciamis Regency, Kuningan Regency, Majalengka Regency, Sumedang Regency, Bogor City, Depok, Tasikmalaya City, and Banjar City, which experience high rainfall but relatively low horticultural production. This indicates underutilized potential requiring improved land management. Cluster 3 comprises Cirebon Regency, Indramayu Regency, Subang Regency, Purwakarta Regency, Karawang Regency, Bekasi Regency, Cirebon City, Bekasi City, and Cimahi City, dominated by higher temperatures and lower rainfall. These regions require irrigation management and heat-resistant crop varieties.

Tabel 2: Clustering Results

No	Regency/City	Cluster
1	Bogor Regency	2
2	Sukabumi Regency	2
3	Cianjur Regency	1
4	Bandung Regency	1
5	Garut Regency	1
6	Tasikmalaya Regency	2
7	Ciamis Regency	2
8	Kuningan Regency	2
9	Cirebon Regency	3
10	Majalengka Regency	2
11	Sumedang Regency	2
12	Indramayu Regency	3
13	Subang Regency	3
14	Purwakarta Regency	3
15	Karawang Regency	3
16	Bekasi Regency	3
17	Bandung Barat Regency	1
18	Pangandaran Regency	2
19	Bogor City	2
20	Sukabumi City	1
21	Bandung City	1
22	Cirebon City	3
23	Bekasi City	3
24	Depok City	2
25	Cimahi City	3
26	Tasikmalaya City	2
27	Banjar City	1

The findings suggest different strategies for each cluster. Cluster 1 needs post-harvest infrastructure and better market access to maintain high production levels. Cluster 2 requires improved land optimization through agricultural technology and water management. Cluster 3 should prioritize drought-tolerant crops and sustainable irrigation systems. Policymakers can use these results to design data-driven strategies that support sustainable horticultural development in West Java Province.

5. Conclussion

This study successfully classified 27 districts and cities in West Java Province based on horticultural indicators using the K-Means method. The optimal result was obtained with three clusters, validated by the highest Silhouette Index score of 0.37, indicating a weak but meaningful clustering structure. The results highlight distinct regional characteristics in horticultural production, temperature, and rainfall, providing valuable insights for regional planning.

Cluster 1 comprises areas with high production and moderate rainfall; Cluster 2 includes regions with high rainfall but relatively low production; and Cluster 3 consists of areas with high temperatures and low rainfall. Overall, the application of K-Means clustering, combined with standardization and Silhouette Index validation, demonstrates a practical and systematic approach to identifying spatial patterns in horticulture. These findings are expected to assist policymakers in formulating targeted strategies that enhance productivity, improve farmer welfare, and support sustainable food security. Future research should incorporate additional indicators, such as soil quality and technology adoption, to improve cluster validity and provide more comprehensive guidance for regional agricultural planning.

Acknowledgments

The authors would like to express their gratitude to Mr. Gerry Sastro and Mr. Tabah Heri Setiawan for their guidance and supervision. Special thanks are also extended to the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Pamulang, for academic support, as well as to BPS and World Climate for providing the data used in this study.

References

Canton, H. (2021). Food and agriculture organization of the United Nations—FAO. In *The Europa directory of international organizations 2021* (pp. 297-305). Routledge.

Christiani, N. V. (2024). Analysis of potential rice production areas in East Nusa Tenggara using K-Means clustering. *Applied Statistics Journal*, *4*(1), 1–11. https://jstar.id/ojs/index.php/JSTAR/article/view/60

Faisal, S. A., & Rifai, N. A. K. (2023). Application of hierarchical clustering multiscale bootstrap for human development index indicators in West Java, 2021. *Bandung Conference Series: Statistics*, *3*(1). https://doi.org/10.29313/bcss.v3i1.6327

Hendrani, Y., Nugraheni, S., & Karliya, N. (2022). Technical efficiency of paddy farming in West Java: a combination of synthetic and organic fertilisers versus conventional farming. *Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)*, *123*(1), 51-62.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons. https://doi.org/10.2307/2532178

Khairiyakh, R. U., & Mulyo, J. H. (2015). Contribution of agricultural sector and sub sectors on Indonesian economy. *Ilmu Pertanian (Agricultural Science)*, *18*(3), 150-159.

Lianita, E., Pratama, A., & Ulva, A. F. (2023). Application of K-Means clustering for mapping vegetable productivity using a GIS-based system in North Sumatra Province. *Justin: Journal of Information Systems and Technology*, *12*(2), 232–238. https://jurnal.untan.ac.id/index.php/justin/article/view/72934/75676600899

Muhima, R. R., Kurniawan, M., Wardhana, S. R., Yudhana, A., Sunardi, Rahmawati, W. M., & Yuliastuti, G. E. (2021). *Clustering algorithms: Concepts, manual calculations, and programming*. Andi Publisher.

Prasetyo, E. (2014). *Data mining: Transforming data into information using Matlab*. Andi Publisher.

Ruslan, K. (2021). Food and horticulture crop productivity in Indonesia (No. 37). Policy Paper.

Saha, S., Kumar, U., Saha, T. N., & Shree, K. (2021). Scope and importance of horticultural crops. In Biointensive Integrated Pest Management for Horticultural Crops (pp. 1-11). CRC Press.

Simatupang, P., & Peter Timmer, C. (2008). Indonesian rice production: policies and realities. Bulletin of Indonesian Economic Studies, 44(1), 65-80.

Singh, A. K., Pathak, M., Joshi, M. D., Kumar, S., Kashyap, S., & Hasan, W. (2024). The role of agriculture in poverty alleviation and rural development. *Journal of Scientific Research and Reports*, *30*(8), 529-549.

Tadi, M., & Ningsih, B. A. (2023). Poverty clustering of districts/cities in Banten Province using the K-Means method. Jurnal Lebesgue: Mathematics and Statistics Education, 4(1), 374–385. https://doi.org/10.46306/lb.v4i1.224

Timmer, P. (2004). Food security in Indonesia: current challenges and the long-run outlook. Center for global development working paper, (48).