Sistemasi: Jurnal Sistem Informasi ISSN:2302-8149
Volume 15, Nomor 1, 2026: 1-10 e-ISSN:2540-9719

Identification of Rice Production Clusters in Central Java Province with K-Means Technique

[1]Yudi Wahyu Wibowo*, [2] Ihsan Cahyo U, [3]Widi Widayat, [4]Juanita Tria S A, [5]As'ad Nirot A
[1,2,3,4,5]Program Studi Teknik Informatika, Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta
[1,2,3,4,5]Jl. A. Yani, Pabelan, Kabupaten Sukoharjo, Jawa Tengah
*e-mail: *yww599@ums.ac.id*

(***received***: 19 May 2025, ***revised***: 21 November 2025, ***accepted***: 22 November 2025)

Abstract

Central Java Province stands as a pivotal agricultural region in Indonesia, characterized by high productivity in paddy and rice cultivation. However, substantial production disparities persist across districts, attributed to varying geographical conditions, agricultural infrastructure, and farmers' cropping patterns. Consequently, the identification of production clusters is imperative for elucidating production patterns and formulating targeted policy interventions. This study aims to classify districts in Central Java based on production metrics using the K-Means Clustering technique implemented in the R statistical environment. Production data across various regions were analyzed to determine optimal clustering patterns. The clustering analysis stratified the study area into four distinct agricultural typologies: optimal performance zones (Cluster 3, n=2), land-based volume producers (Cluster 1, n=7), small-scale efficient producers (Cluster 4, n=15), and priority intervention areas (Cluster 2, n=11). These findings underscore the necessity for differentiated policy strategies addressing the disparities in efficiency and production scales.

Keywords: central java, k-means clustering, r, rice production

1 Introduction

Agriculture is a crucial sector in Indonesia's economy, particularly in ensuring national food security. Rice, as a primary commodity, plays a significant role in maintaining food resilience for the Indonesian population. Food security in Indonesia is a strategic issue that continues to receive attention, given the country's large population and high dependence on the agricultural sector[1] [2]. Factors such as population growth, climate change, land-use conversion, and fluctuations in global food prices pose major challenges in maintaining food availability, accessibility, and stability [3]. The government has implemented various policies, including local food diversification, strengthening the logistics system, and providing subsidies for fertilizers and seeds to boost domestic production. However, there remains a gap between surplus and deficit regions, leading to uneven food distribution and reliance on imports of key commodities such as wheat and soybeans [4] [5]. Therefore, precision agriculture technology, farmer institutional empowerment, and the optimization of sustainable agricultural land are crucial solutions to address these challenges. On the other hand, food security is not only dependent on food availability but also on economic access and affordability for society[6] [1]. Poverty and income inequality remain key factors limiting household purchasing power, particularly in rural areas and poverty-stricken regions. Therefore, strengthening social assistance programs, such as the Non-Cash Food Assistance (BPNT) program and food price subsidies for vulnerable groups, is essential to ensuring more equitable food access [7]. Additionally, education on more diverse and nutritious dietary patterns is necessary to reduce dependence on rice as the staple food. With a holistic approach encompassing food production, distribution, and consumption, Indonesia can enhance its food security and achieve sustainable self-sufficiency.

Central Java Province is one of Indonesia's key rice-producing regions, according to data from BPS (Statistics Indonesia), Central Java was the second-largest national rice producer in 2023, contributing approximately 9.06 million tons of dry unhusked rice (GKG), which accounts for around 16.78% of Indonesia's total national production. This substantial contribution underscores its critical role in the national rice supply. However, the distribution of rice production across Central Java is uneven. Geographic differences, climate variations, and access to agricultural technology leads to considerable disparities in production yields across districts and cities. Identifying regions with similar or distinct production characteristics is essential for understanding the production dynamics on the ground. This information can be utilized by the government and stakeholders to formulate more effective policies tailored to the needs of each region. One of the methods to analyze the rice production patterns in Central Java is through cluster analysis. The K-Means clustering technique is an effective statistical method for grouping data based on similar characteristics, which, in this context, can be applied to categorize regions according to their rice production patterns [8] [9]. By implementing this technique, it is expected that distinct clusters of rice-producing regions can be identified, providing a basis for designing more targeted strategies to enhance production. These strategies may include resource allocation, technological assistance, and the distribution of fertilizers and other agricultural inputs. Furthermore, clustering will help optimize regional potential and reduce production disparities among districts and cities in Central Java Province. Based on this background, the research aims to address the following problem statements: (1) What are the distribution patterns of rice production across the districts and cities in Central Java Province? (2) How can the K-Means clustering technique be utilized to identify clusters of rice-producing regions in Central Java?

Accordingly, the main objectives of this study are: (1) to analyze and map the distribution patterns of rice production across districts and cities in Central Java, and (2) to implement the K-Means algorithm to group these regions into distinct clusters based on their production characteristics. The expected benefit of this research is to provide a data-driven grouping that serves as a decision-making tool for the provincial government and related stakeholders. This will aid in formulating more effective, targeted policies for resource allocation (e.g., fertilizer subsidies, irrigation development) and technological assistance, tailored to the specific needs of each cluster. The novelty of this study lies in its application of K-Means to recent, comprehensive production data for all districts/cities in Central Java, offering an updated and specific map for regional agricultural planning aimed at reducing production disparities, which has not been the focus of similar previous research [10].

2 Literature Review

K-Means is one of the most widely used clustering algorithms in data analysis, particularly for grouping unlabeled data. This algorithm partitions a dataset into k clusters based on the proximity between data points and cluster centroids [11]. K-Means employs an iterative approach to minimize intra-cluster variability by continuously updating the centroid positions until convergence. This technique has been applied in various fields, including customer segmentation in marketing image analysis and pattern recognition in social and economic studies. Although effective for handling large datasets, K-Means has several limitations, such as sensitivity to the initial number of clusters and a tendency to converge to local optima. To address these challenges, numerous studies have proposed enhancements to the K-Means method. One commonly adopted approach is the K-Means++ algorithm, which improves centroid initialization to produce more stable and optimal clusters. Additionally, hybrid methods combining K-Means with hierarchical or density-based clustering have been developed to enhance clustering accuracy in complex datasets. [12]The integration of K-Means with parallel computing and machine learning techniques has further expanded its applications in the era of big data. As a result, K-Means remains a highly relevant technique in data exploration, particularly for high-dimensional clustering analysis that demands both efficiency and interpretability.

However, despite these advancements, several critical gaps remain, particularly in the application of K-Means to complex, real-world datasets like regional agricultural production. Firstly, previous studies often fail to adequately address the challenge of clustering mixed data types [13] . Standard K-Means relies on Euclidean distance, which is only suitable for numerical data (e.g., production tonnage, land area). It cannot natively handle the categorical data often present in agricultural analysis (e.g., district names, soil types, main crop varieties). This forces researchers to either discard valuable categorical information or use less-than-ideal workarounds, whereas more appropriate methods like K-Prototypes, which can handle mixed data, are often overlooked [14].
Secondly, most applications of K-Means for regional or geospatial data, such as grouping districts by production, do not incorporate spatial constraints [15]. The algorithm may group a district in the far north of a province with one in the far south simply because their production values are similar. This creates clusters that are geographically fragmented and impractical for policy implementation, which often requires geographically contiguous zones for efficient resource management [16]. Research has

ISSN:2302-8149
e-ISSN:2540-9719

not yet fully integrated spatial contiguity constraints directly into the K-Means clustering process for agricultural zoning.

Finally, while sensitivity to outliers is a known limitation, many studies applying K-Means to agricultural data do not explicitly address this. Agricultural production data is frequently skewed by extreme events like droughts, floods, or pest attacks in specific areas [15]. These outliers can disproportionately pull cluster centroids, leading to distorted and unrepresentative groupings. Consequently, there is a need for research that not only applies K-Means but also integrates robust outlier detection or uses more resilient variations of the algorithm to ensure the resulting clusters reflect typical production patterns rather than anomalies.

3 Research Method

This study employs a cluster analysis approach using the K-Means clustering method to identify regional groupings based on rice and paddy production characteristics in Central Java Province. This method was selected due to its capability to classify data into clusters with similar characteristics based on specific variables. In this research, the analysis is conducted using R software. The research flow is depicted in Figure 1 Research flow

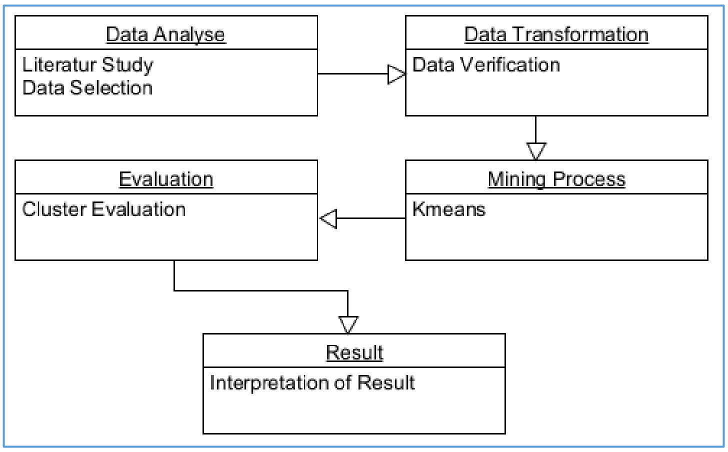

Figure 1 Research flow

Based on the research steps from **Figure 1 Research flow** is as follow:

1. Data Analyst in this stage, it is carried out to study references in the form of journals, books and other articles related to research. The references for this research include journals that explore topics related to data mining, big data, and the K-Means algorithm. Additionally, books and articles that discuss K-Means clustering analysis are also utilized as key sources of information. Furthermore, the selection of data plays a crucial role in this study, as it serves as the foundation for the research materials. Careful consideration is given to ensure that the chosen data aligns with the objectives of the study.
2. Data Transformation The data transformation process involves preparing data so that it meets the requirements of the k-means algorithm. This step ensures that the data is clean and suitable for clustering analysis. One essential aspect of this process is handling incomplete or missing data. This can be done by removing incomplete records or imputing missing values using methods such as the mean, median, mode, or interpolation. Additionally, data duplication may be addressed to maintain data integrity [17].
3. Data mining is an important process in data analysis that aims to extract hidden knowledge or patterns from large and complex datasets. To cluster regions based on their production characteristics, the K-Means clustering technique is used. [12]This clustering process includes several stages, namely: - Data Normalization: Since the variables used have different scales (for example, ton for production and hectares for land area), normalization is performed so that all variables are on the same scale. - Determination of Number of Clusters: One of the challenges in the K-Means method is determining the optimal number of clusters (k).[18] For this, the Elbow or Silhouette Score method is used to find the right number of clusters, that is where adding more clusters no longer provides significant improvement in clustering quality [19]. K-Means Iteration:

ISSN:2302-8149
e-ISSN:2540-9719

The K-Means algorithm works by randomly initializing the cluster centers, then iterating to move the cluster centers until convergence, that is when the change in cluster position is no longer significant[12].

4. Evaluation The evaluation stage is very important to ensure that the clusters formed truly represent the patterns in the data well and can be interpreted. Evaluation methods such as Elbow, Silhouette Score, cluster visualization, and stability analysis can ensure that the K-Means clustering results are accurate. A good evaluation also helps to provide more appropriate policy recommendations based on the clustering results.
5. Result. This study successfully identified several clusters of paddy and rice production in Central Java Province using the K-Means technique. The clustering results provide a clear picture of the differences in regional characteristics in terms of productivity, land area, and total production, which can be a reference for the government in designing evidence-based policies to improve the efficiency and productivity of the rice farming sector in Central Java.

4 Results and Analysis

Based on **Table 1** regarding Agriculture Data, the data used is data on the area, production and productivity of rice in Central Java in 2024 [20]. Indicator data included are only those related to land area, rice production and rice productivity. The indicators are: Land area (Area), which is the average area of paddy fields in hectares. Rice production produced (Production), which is the number of harvests in tons. Rice production (Productivity), which is the amount of rice produced from the number of harvests divided by the area of land (ku/ha). The purpose of cluster analysis is to cluster regions based on productivity indicators that include area, production, and productivity indicators. These indicators can be used to determine the quality of rice productivity in Central Java.

Table 1 Agriculture data

No	City	Harvest Area	Production	Productivity
1	Cilacap	119730,72	772489	64,52
2	Banyumas	51330,54	280224	54,59
3	Purbalingga	26110,28	140161	53,68
4	Banjarnegara	18552,34	96637,5	52,09
5	Kebumen	75547,03	404318	53,52
6	Purworejo	51160,94	287721	56,24
7	Wonosobo	12578,2	60651,9	48,22
8	Magelang	31381,63	160695	51,21
9	Boyolali	49819,29	277038	55,61
10	Klaten	64377,47	353911	54,97
11	Sukoharjo	49628,07	348736	70,27
12	Wonogiri	66897,03	376431	56,27
13	Karanganyar	42874,47	266578	62,18
14	Sragen	109077,98	641988	58,86
15	Grobogan	129630,82	678350	52,33
16	Blora	98784,77	472869	47,87
17	Rembang	38263,49	215629	56,35
18	Pati	93852,59	508150	54,14
19	Kudus	29090,76	162001	55,69
20	Jepara	38029,74	187317	49,26
21	Demak	99263,51	564962	56,92

ISSN:2302-8149
e-ISSN:2540-9719

22	Semarang	26566,2	143194	53,9
23	Temanggung	8009,97	46499,4	58,05
24	Kendal	29566,56	158931	53,75
25	Batang	26690,07	146350	54,83
26	Pekalongan	38115,12	169539	44,48
27	Pemalang	73832,38	384839	52,12
28	Tegal	59835,78	322355	53,87
29	Brebes	78814,2	424363	53,84
30	Magelang	123,5	646,43	52,34
31	Surakarta	27,09	157,57	58,17
32	Salatiga	582,55	3501,04	60,1
33	Semarang	2765,75	15925,6	57,58
34	Pekalongan	1305,89	7766,8	59,48
35	Tegal	544,5	3181,05	58,42

Stages Before Cluster Analysis Data Normalization Because the data used has different units and ranges, it is necessary to standardize the data first before entering cluster analysis. To standardize data into standard normal in R, you can use the scale() function [21] [22].

```
data_standardized <- round(scale(beras[,2:4]),3)
```

Determining the number of clusters. In the k-means method the number of clusters is determined by the user. Therefore, it is necessary to find the optimum number of clusters that can group objects well (Please note that this method is relatively subjective). One of the methods used is the *Elbow Plot*. Elbow Plot is a plot between the number of clusters and the total within-cluster variation. The number of clusters chosen is the “elbow” or the point where there is a sharp drop before that point and a less sharp drop after that point. This is because increasing the number of clusters does not have much effect on the within-cluster variation.

The determination of the optimal number of clusters (k) was conducted using the Elbow Method within the RStudio environment. Specifically, the `fviz_nbclust()` function was applied to the standardized dataset using the method = `'wss'` argument. The resulting plot was analyzed to identify the inflection point—commonly referred to as the 'elbow'—where the reduction in the Within Cluster Sum of Squares (WSS) begins to plateau significantly; this point was subsequently established as the optimal k value for the analysis. The code is

```
fviz_nbclust(data_scaled, kmeans, method = "wss") + geom_vline(xintercept
= 4, linetype = 2) + labs(subtitle = "menentukan jumlah klaster ")
```

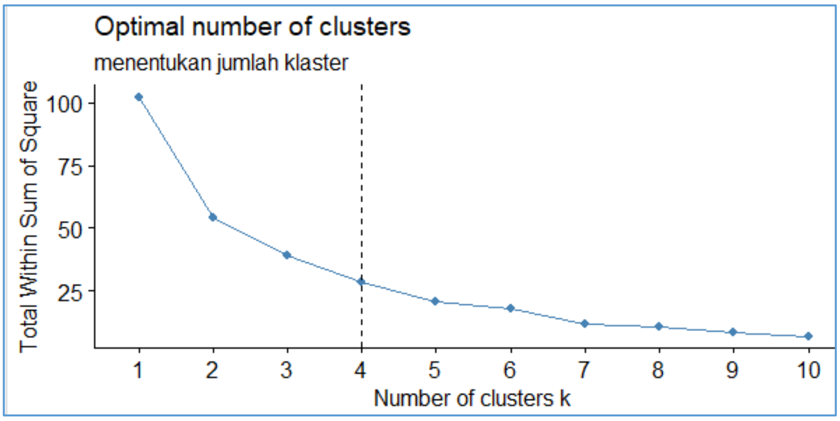

Figure 2 Elbow plot result

ISSN:2302-8149
e-ISSN:2540-9719

As depicted in the Elbow plot (**Figure 2 Elbow plot result**), the curve suggests that the optimal number of clusters is four (k=4). This parameter was subsequently applied in the K-Means clustering analysis using RStudio.

The K-means clustering algorithm was subsequently executed in the RStudio environment using the following command:

```
set.seed(123)
> kmeans.proc <- kmeans(data_scaled, centers = 4, nstart = 25)
```

```
> kmeans.proc
K-means clustering with 4 clusters of sizes 7, 11, 2, 15

Cluster means:
        Luas   Produksi Produktivitas
1 -1.0615527 -1.0120506     0.8023224
2  1.0751753  0.9956500    -0.2598723
3  1.0294519  1.4477637     2.5299347
4 -0.4303309 -0.4508882    -0.5211688
```

Table 2 Cluster profiles of agricultural regions based on standardized means

Cluster	Count (n)	Harvested Area	Production	Productivity	Cluster Typology & Interpretation
Cluster 1	7	-1.061	-1.012	+0.802	High Efficiency: Small scale and low production volume, but achieves high productivity (indicative of intensive/urban farming).
Cluster 2	11	1.075	+0.996	-0.259	Extensive Farming: Large land area and high production volume, but productivity remains slightly below average.
Cluster 3	2	1.029	1.447	**2.529**	Superior Performance: The top-performing regions featuring large areas, high output, and the highest efficiency.
Cluster 4	15	-0.430	-0.450	-0.521	Low Performance: Regions with medium-to-low scale (area/production) accompanied by relatively low productivity.

Refers to Table 2 Cluster profiles of agricultural regions based on standardized means, Interpretation of Cluster Characteristics. The four identified clusters offer specific insights for policy intervention:

The K-Means clustering analysis stratified the 35 studied regions into four distinct clusters based on harvested area, production volume, and productivity dimensions. The standardized centroid values reveal unique typologies for each group:

Cluster 1 (n=7) represents regions characterized by limited land area (-1.06) and low production volume (-1.01), yet exhibiting significantly above-average productivity (+0.80). This pattern strongly reflects urban farming characteristics or intensive agricultural zones where land efficiency is maximized.

ISSN:2302-8149
e-ISSN:2540-9719

In contrast, Cluster 2 (n=11) is defined as an extensive production hub, marked by positive deviations in land area (+1.07) and production (+0.99), despite exhibiting slightly below-average productivity (-0.25).

Notably, Cluster 3 (n=2) emerges as the superior group ('high-performing outliers'). Regions in this cluster possess not only substantial land area (+1.02) and production capacity (+1.44) but also the highest productivity score (+2.52), indicating highly effective agricultural optimization.

Finally, Cluster 4 (n=15) comprises the majority of regions showing below-average performance across all three variables, suggesting a critical need for targeted policy interventions to enhance agricultural capacity..

Cluster analysis in R can be performed using the kmeans() function and standardized data.[23] [24] One crucial aspect to consider when conducting cluster analysis using the K-Means algorithm is the initialization of the initial centroids. This is particularly important because the K-Means algorithm begins by randomly placing (or initializing) the initial centroids within the dataset. Consequently, the clustering results may vary across different runs if only a single initialization is used. To address this issue and obtain an optimal clustering outcome, the algorithm should be executed multiple times with different initial centroid values. In the kmeans() function, the parameter nstart specifies the number of times the algorithm should initialize the centroids. In this example, the nstart parameter is set to 25, meaning that the algorithm will generate 25 different initial centroid configurations [25]. Subsequently, the K-Means algorithm will select the best clustering result by identifying the initialization that yields the smallest total within-cluster variation among the 25 trials. A lower total within-cluster variation indicates that, in general, the clusters formed consist of observations that are more homogeneous or similar to one another.

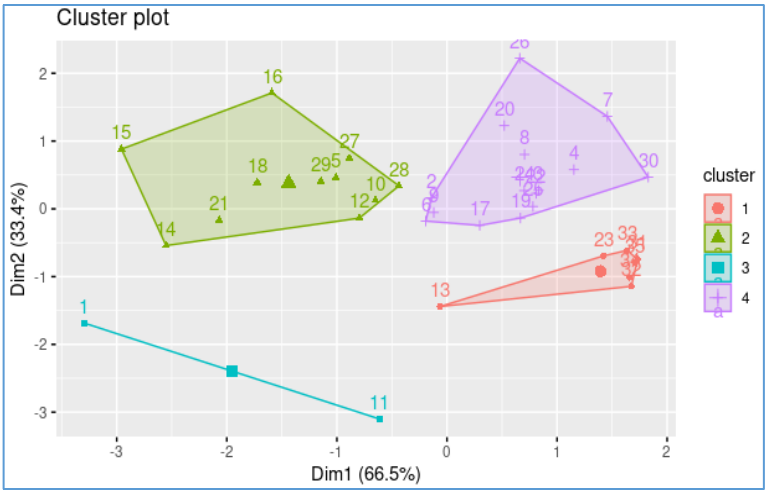

Figure 3 Cluster plot visualizing with R

The clustering results provide a clear picture of the differences in regional characteristics in terms of productivity, land area, and total production. Based on the visualized Cluster Plot (**Figure 3 Cluster plot visualizing with R**) the cluster interpretation:

Cluster 1 (Red): Urban & High-Efficiency Zones (n=7) Located on the negative side of Dim1, this cluster aggregates regions with limited land scale but high spatial efficiency. It includes all administrative cities (Kota Magelang, Surakarta, Salatiga, Semarang, Pekalongan, Tegal) and Kab. Kudus. Despite contributing minimally to total provincial volume, these regions exhibit high land-use intensity, characterizing them as urban or peri-urban agricultural systems where productivity per hectare is maximized.

The validity of these clustering results confirms the methodological robustness in handling the heterogeneity of the spatial units. Despite the significant physical scale disparity between administrative cities (small urban units) and regencies (large agricultural units), the application of data standardization (Z-score normalization) proved effective in mitigating magnitude bias. Conceptually, the formation of Cluster 1 demonstrates that the inclusion of urban areas constitutes a distinct agricultural typology—representing intensive urban farming—rather than a disruptive outlier. Excluding these cities would result in the loss of a critical dimension regarding high-efficiency land

ISSN:2302-8149
e-ISSN:2540-9719

utilization (where Zprod>0.80), which remains an integral part of the holistic agricultural landscape of Central Java.

Cluster 2 (Green): Extensive Food Granaries (n=11) Positioned on the far right (positive Dim1), this group represents the province's agricultural backbone, characterized by extensive harvested areas and massive production volumes. Members include major producers like Kab. Grobogan, Cilacap, Pati, Sragen, Demak, Banyumas, Brebes, Kebumen, Purworejo, Blora, and Pemalang. While dominant in scale, their lower placement on Dim2 suggests productivity levels are average or slightly below optimal compared to the high-performing cluster.

Cluster 3 (Blue): Superior Performance Outliers (n=2) Occupying the unique top-right quadrant (Positive Dim1 and High Positive Dim2), Kab. Klaten and Kab. Sukoharjo emerge as the "best-practice" group. They are distinct outliers that successfully combine large-scale operations with the highest productivity rates in the province (Z>2.5), separating them significantly from the extensive granaries in Cluster 2.

Cluster 4 (Purple): Low-Input/Subsistence Zones (n=15) Clustered near the center-left origin, this group comprises regions with below-average performance across scale and productivity dimensions. It encompasses Kab. Purbalingga, Banjarnegara, Wonosobo, Magelang, Boyolali, Wonogiri, Karanganyar, Rembang, Jepara, Semarang, Temanggung, Kendal, Batang, Pekalongan, and Tegal. The clustering suggests these regions face geographical constraints (e.g., highland terrain in Wonosobo/Temanggung) or transitional land-use patterns that limit agricultural expansion.

Through the implementation of K-Means Clustering analysis on rice and paddy production in Central Java Province using the R programming language, this study has successfully classified regions into four distinct clusters based on their production characteristics. The first cluster comprises seven regions characterized by high production levels, while the second cluster consists of eleven regions with moderate production. Meanwhile, the third cluster includes two regions exhibiting exceptionally high production levels, and the fourth cluster encompasses fifteen regions with relatively lower production output.

This study's approach provides crucial advantages compared to existing literature: Policy Granularity: The key strength is the focus on the district/city level within Central Java. Prior national or provincial-level clustering studies often masked the significant variations identified here. This fine-grained analysis is indispensable for designing evidence-based policies, allowing the government to tailor resource allocation, fertilizer distribution, and technological application precisely to the needs of each cluster, a level of detail affirmed as critical for effective agricultural planning in Indonesia [10]. Methodological Robustness: By setting the nstart parameter to 25 , the research effectively mitigates the inherent weakness of K-Means—its sensitivity to initial centroid placement . This methodological rigor ensures that the resulting clusters are stable and represent an optimal partitioning of the data, thereby increasing the reliability of the policy recommendations derived [26]. Contemporary Relevance: The use of 2024 data ensures that the production patterns identified reflect current agricultural realities, accounting for recent land-use shifts or climate impacts. This contemporaneity makes the findings immediately useful for current governmental planning.[16]

5 Conclusion

This study successfully mapped spatial disparities in rice productivity across Central Java Province by applying the K-Means Clustering algorithm. The primary objective of identifying production patterns to formulate effective policy strategies was achieved by uncovering significant regional heterogeneity. The statistical analysis stratified the 35 studied regions into four distinct agricultural typologies based on harvested area, production volume, and productivity dimensions. Empirical findings demonstrate a polarization of regional characteristics: **Cluster 3** (n=2) as the "Superior Performance Zone" (Klaten-Sukoharjo), achieving the highest efficiency (Z>2.5); **Cluster 2** (n=11) as "Extensive Food Granaries," dominating total output through massive land scale despite stagnant productivity; **Cluster 1** (n=7) representing efficiency within constrained lands (urban characteristics); and **Cluster 4** (n=15) as subsistence zones with below-average performance in both scale and productivity. The managerial implication of these findings confirms that a "one-size-fits-all" policy approach is no longer relevant. Differentiated interventions are required, ranging from technological intensification for Cluster 2 to the revitalization of basic inputs for Cluster 4. However, this study acknowledges specific limitations. First, the clustering analysis is cross-sectional, representing a single time period, and thus does not capture seasonal dynamics or long-term production trends. Second, the variables were limited to production outputs (area, production, productivity) without integrating external causal variables mentioned in the background, such as irrigation infrastructure conditions, climate anomalies, or soil characteristics. Consequently, future research is recommended to: (1) apply spatio-temporal analysis using time-series data to validate cluster stability over time; and (2) enrich the model by incorporating environmental and infrastructural variables (such as drought indices or percentage of technical irrigation) to provide a more holistic and precise determinant analysis for policymakers.

Acknowledgement

The author wishes to express profound gratitude to all parties who have provided invaluable support, guidance, and contributions throughout the completion of this research. Special appreciation is extended to the research team for their insightful direction and constructive feedback at every stage of the study. Sincere gratitude is also conveyed to RISETMU for its generous support and funding, which have greatly facilitated both the research process and its subsequent publication. Finally, heartfelt appreciation is dedicated to the author's family and all individuals who have offered unwavering moral support and encouragement. It is the author’s sincere hope that this research will contribute meaningfully to the advancement of scientific knowledge..

Reference

[1] A. N. Ayu Mutia, I. Nurlinda, and N. Astriani, “Pengaturan Pembangunan *Food Estate* pada Kawasan Hutan untuk Mewujudkan Ketahanan Pangan di Indonesia,” *Bina Hukum Lingkungan*, Vol. 6, No. 2, pp. 224–240, Mar. 2022, DOI: 10.24970/bhl.v6i2.259.

[2] I. Ahmadian, A. Yustiati, and D. Y. Andriani, “Produktivitas Budidaya Sistem Mina Padi untuk Meningkatkan Ketahanan Pangan di Indonesia: A Review.”

[3] P. Pratama Siregar and Z. Almaida Siregar, “Resolusi : Rekayasa Teknik Informatika dan Informasi Penerapan Metode K-Means dalam Mengelompokkan Persebaran Lahan Kritis di Indonesia berdasarkan Provinsi,” Vol. 2, No. 4, pp. 145–151, 2022, [Online]. Available: https://djournals.com/resolusi

[4] N. Salsabila and Hendy Tannady, “Manajemen Distribusi Komoditas Harga Bahan Pangan di Indonesia dengan Metode *Least Cost*,” Jan. 2021, DOI: 10.31219/osf.io/n37rg.

[5] A. Adiyanto and Y. Arie Wijaya, “Penerapan *Algoritma K-Means* pada Pengelompokan Data Set Bahan Pangan Indonesia Tahun 2022-2023,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 7, No. 2, pp. 1344–1350, Sep. 2023, DOI: 10.36040/jati.v7i2.6849.

[6] A. Rasman, E. Sinta Theresia, and dan M. Fadel Aginda, “Analisis Implementasi Program *Food Estate* sebagai Solusi Ketahanan Pangan Indonesia,” *Holistic: Journal of Tropical Agriculture Sciences Riset*, Vol. 1, No. 1, pp. 36–68, 2023, DOI: 10.61511/hjtas.v.

[7] S. Komunikasi Pemerintah dalam Mensosialisasikan Bantuan Pangan Non Tunai di Desa Simpangan Kecamatan Cikarang Utara Fihris Sa and E. Priyanti, “Strategi Komunikasi Pemerintah dalam Mensosialisasikan Bantuan Pangan Non Tunai (BPNT) di Desa Simpangan Kecamatan Cikarang Utara,” *j-innovative.org*, Accessed: Feb. 16, 2025. [Online]. Available: http://j-innovative.org/index.php/Innovative/article/view/11409

[8] A. Larasati, … R. M.-J. T. I., and undefined 2021, “*Utilizing Elbow Method for Text Clustering Optimization in Analyzing Social Media Marketing Content of Indonesian e-Commerce*,” *jurnalindustri.petra.ac.id*, Accessed: Feb. 16, 2025. [Online]. Available: https://jurnalindustri.petra.ac.id/index.php/ind/article/view/23840

[9] M. M. abdoel Wahid, “*Determining the Location of RMU, using K-Means Clustering, Evaluate the Location of Existing RMU, using R-Programming*,” *Journal of Informatics and Telecommunication Engineering*, Vol. 6, No. 1, pp. 10–17, Jul. 2022, DOI: 10.31289/jite.v6i1.6126.

[10] T. Tuslaela, R. Rusdiansyah, H. Supendar, and N. Suharyanti, “*Implementation of K-Means Clustering in Food Security by Regency in East Java Province in 2022*,” *Sinkron*, Vol. 9, No. 1, pp. 54–60, Jan. 2024, DOI: 10.33395/SINKRON.V9I1.13169.

ISSN:2302-8149
e-ISSN:2540-9719

[11] D. Steinley, "*K-Means Clustering: A Half-Century Synthesis*," *British Journal of Mathematical and Statistical Psychology*, Vol. 59, No. 1, pp. 1–34, May 2006, DOI: 10.1348/000711005X48266.

[12] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "*International Journal of Advances in Scientific Research and Engineering (IJASRE) Comprehensive Review of K-Means Clustering Algorithms*," 2021, DOI: 10.31695/IJASRE.2021.34050.

[13] A. Ahmad and L. Dey, "*A K-Mean Clustering Algorithm for Mixed Numeric and Categorical Data*," *Data Knowl Eng*, Vol. 63, No. 2, pp. 503–527, Nov. 2007, DOI: 10.1016/J.DATAK.2007.03.016.

[14] "*The Ultimate Guide for Clustering Mixed Data | by Eoghan Keany | Analytics Vidhya | Medium*." Accessed: Nov. 05, 2025. [Online]. Available: https://medium.com/analytics-vidhya/the-ultimate-guide-for-clustering-mixed-data-1eefa0b4743b

[15] F. Zhu, C. Zhu, Z. Fang, W. Lu, and J. Pan, "*Using Constrained K-Means Clustering for Soil Texture Mapping with Limited Soil Samples*," *Agronomy 2025, Vol. 15, Page 1220*, Vol. 15, No. 5, p. 1220, May 2025, DOI: 10.3390/AGRONOMY15051220.

[16] I. Mujahidin and S. H. Hasanah, "*Comparative Analysis of K-Means, K-Medoids, and Fuzzy C-Means for Clustering Provinces in Indonesia based on Rice Production in 2024*," *Jurnal Gaussian*, Vol. 14, No. 2, pp. 356–365, Oct. 2025, DOI: 10.14710/J.GAUSS.14.2.356-365.

[17] G. Bellinger, D. Castro, and A. Mills, "*Data, Information, Knowledge, and Wisdom*." [Online]. Available: http://outsights.com/systems/dikw/dikw.htm

[18] A. Fahim, "*K and Starting Means for K-Means Algorithm*," *J Comput Sci*, Vol. 55, p. 101445, Oct. 2021, DOI: 10.1016/J.JOCS.2021.101445.

[19] P. M. Hasugian, B. Sinaga, J. Manurung, and S. A. Al Hashim, "*Best Cluster Optimization with Combination of K-Means Algorithm and Elbow Method Towards Rice Production Status Determination*," *International Journal of Artificial Intelligence Research*, Vol. 5, No. 1, Jun. 2021, DOI: 10.29099/ijair.v6i1.232.

[20] "Luas Panen, Produktivitas, dan Produksi Padi Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2024 - Tabel Statistik - Badan Pusat Statistik Provinsi Jawa Tengah." Accessed: May 19, 2025. [Online]. Available: https://jateng.bps.go.id/id/statistics-table/3/WmpaNk1YbGFjR0pOUjBKYWFIQlBSU3MwVHpOVWR6MDkjMyMzMzAw/luas-panen--produktivitas--dan-produksi-padi-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html?year=2024

[21] G. W. Milligan and M. C. Cooper, "*A Study of Standardization of Variables in Cluster Analysis*," *J Classif*, Vol. 5, No. 2, pp. 181–204, Sep. 1988, DOI: 10.1007/BF01897163/METRICS.

[22] "*Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning - Alboukadel Kassambara - Google Books*." Accessed: May 19, 2025. [Online]. Available: https://books.google.co.id/books?hl=en&lr=&id=plEyDwAAQBAJ&oi=fnd&pg=PP2&dq=step+cluster+on+R&ots=xeI_lDjR0z&sig=6Fhz-3j15UGpZ1y92EgnLLqvynA&redir_esc=y#v=onepage&q=step%20cluster%20on%20R&f=false

[23] "*Cluster Analysis of Charitable Organizations of Ukraine using K-Means Technology*," *Revista »Administratie si Management Public« (RAMP)*, No. 37, pp. 117–131, 2021.

[24] R. Scitovski, K. Sabo, F. Martínez-Álvarez, and Š. Ungar, "Cluster Analysis and Applications," *Cluster Analysis and Applications*, pp. 1–271, Jan. 2021, DOI: 10.1007/978-3-030-74552-3/COVER.

[25] D. Gurukumaresan, C. Duraisamy, R. Srinivasan, and V. Vijayan, "*Optimal Solution of Fuzzy Assignment Problem with Centroid Methods*," *Mater Today Proc*, Vol. 37, No. Part 2, pp. 553–555, Jan. 2021, DOI: 10.1016/J.MATPR.2020.05.582.

[26] A. Ahdika, M. D. Kartikasari, S. K. Dini, and I. Ramadhani, "*Diversification of Agricultural Areas in Indonesia using Dynamic Copula Modeling and K-Means Clustering*," *Sains Malays*, Vol. 50, No. 9, pp. 2791–2817, Sep. 2021, DOI: 10.17576/JSM-2021-5009-24.