

Hierarchical Clustering for Rice Planting Season Recommendations in Subak Tabanan

Ni Made Cahyani Dewi, Ahmad Tri Hidayat
Yogyakarta University of Technology, Indonesia

ABSTRACT

This research applies hierarchical clustering to classify rice harvest periods in Subak Tabanan, Bali, using monthly harvest area data from 2020–2024. The objective is to identify seasonal patterns that can guide planting recommendations for local farmers. Data preprocessing involved standardization and transformation into numerical format suitable for clustering. The analysis focused on three clusters representing rainy season, transitional season, and dry season. The results indicate that most months fall within the rainy season cluster, while transitional months and a single dry month were distinctly identified. The silhouette score value shows moderate clustering performance, indicating that hierarchical clustering is capable of distinguishing planting seasons effectively. Visualization through dendrogram and distribution charts supported the identification of cluster groups. This study contributes to agricultural decision support systems, particularly in improving planting strategies and ensuring rice production sustainability in Subak Tabanan.

Keywords: Hierarchical Clustering, Rice Harvest, Planting Season, Subak Tabanan

Corresponding author

Name: Ni Made Cahyani Dewi

Email: cahyanidewi86@gmail.com

INTRODUCTION

Indonesia is widely recognized as an agrarian country, where rice plays a vital role as the staple food for the majority of its population. In Bali, the traditional Subak system is not only an irrigation system but also a socio-cultural institution that manages water distribution and farming practices collectively. Subak in Tabanan Regency, known as the “rice barn of Bali,” has long contributed to the sustainability of rice production. However, farmers often face difficulties in determining the optimal planting season due to unpredictable climate variability, especially the shifting patterns of the rainy and dry seasons. This uncertainty often leads to imbalanced harvest distribution, risk of crop failure, and decreased productivity.

Conventionally, the determination of planting and harvesting schedules in Subak has been based on local experience and farmer intuition. While such traditional practices have cultural and historical value, they may lack the precision required to address the challenges posed by climate change and modern agricultural demands. Consequently, data-driven approaches are increasingly necessary to optimize planting strategies and ensure

sustainable rice production. One such approach is clustering analysis, which has been widely applied in agriculture to identify patterns and groupings within large datasets (Javadi et al., 2022; Guevara-Viejó & Martínez, 2021).

Hierarchical Clustering is employed in this study to classify monthly rice harvest data from 2020–2024 in Subak Tabanan. Unlike other clustering methods such as K-Means, which requires predefined cluster numbers, Hierarchical Clustering generates a hierarchical tree structure (dendrogram) that visualizes the relationships among data points (Sitokonstantinou et al., 2023). This capability makes Hierarchical Clustering particularly suitable for identifying seasonal patterns in agricultural cycles. Previous studies have widely applied clustering for agricultural purposes, including yield monitoring, mapping productivity zones, and crop phenology estimation (Shawon et al., 2024; Rosiana et al., 2025). However, applications of Hierarchical Clustering for seasonal rice harvest classification in Subak Bali remain limited, as most research relies on K-Means clustering (Sanela et al., 2023; Farismana, 2024).

This study aims to classify rice harvest periods into three seasonal clusters: rainy season, transitional season, and dry season by applying Hierarchical Clustering with standardized monthly data input. Cluster quality is evaluated using the silhouette score and supported through visualization such as dendrograms and distribution charts. The findings are expected to provide a more objective seasonal analysis, strengthen data-driven decision-making in Subak management, and contribute to the development of sustainable agricultural planning systems amid increasing climate uncertainty in Bali.

METHOD

The research design adopts a quantitative descriptive method, aiming to identify and interpret seasonal patterns in rice harvest areas in Subak Tabanan, Bali. The dataset consists of monthly rice harvest area records from 2020 to 2024, obtained from official agricultural statistics. Each record represents one observation unit, corresponding to the harvested area per month. This dataset serves as the primary input for clustering analysis to classify seasonal planting patterns.

The workflow of this study is illustrated in Figure 1, which outlines the step-by-step process. The initial stage begins with data collection of rice harvest areas from 2020 to 2024. The collected data is then subjected to data cleaning and preprocessing, where inconsistencies, missing values, and irrelevant information are handled to ensure the dataset is complete and reliable. Following this, the dataset undergoes standardization using *StandardScaler*, which normalizes the range of values so that all features contribute equally to the clustering process. Once standardized, the dataset is transformed into numerical format, making it suitable for machine learning processing.

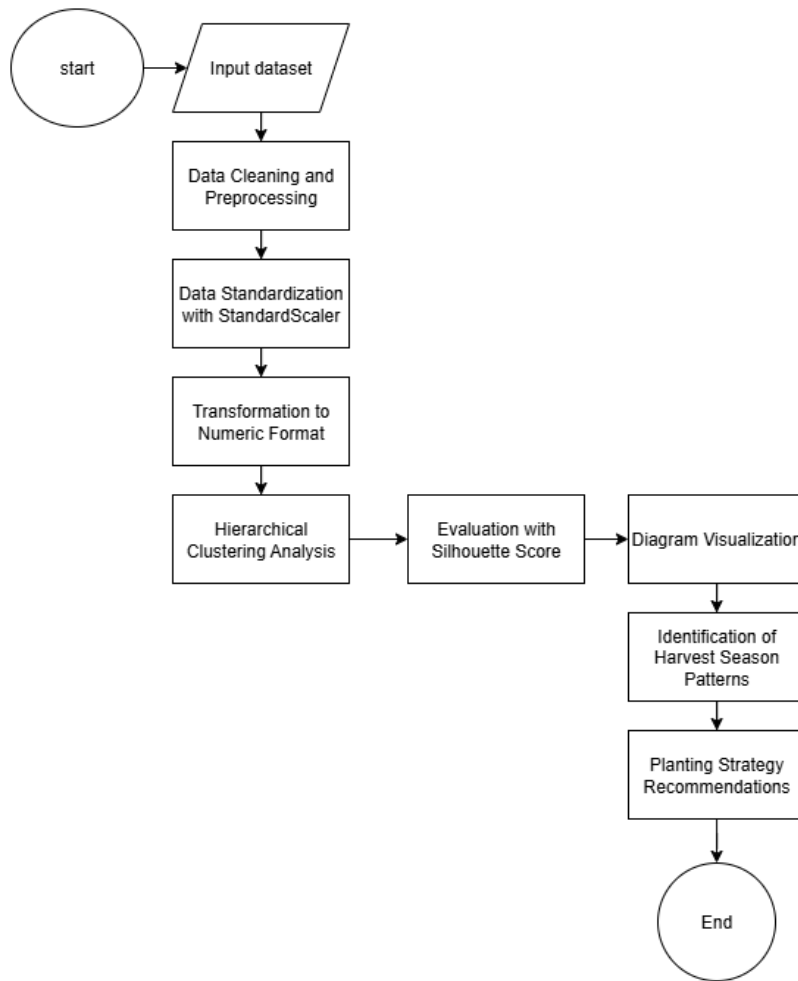


Figure 1. Flowchart of the hierarchical clustering System

After preprocessing, the next stage involves the application of hierarchical clustering analysis using Ward’s linkage method. This technique groups the data into three main clusters representing the rainy season, transitional season, and dry season. The optimal number of clusters is determined based on agricultural domain knowledge and confirmed through cluster evaluation metrics. To validate the clustering quality, the Silhouette Score is applied, providing a measure of how well each data point fits within its assigned cluster compared to other clusters.

The results of clustering are then represented through visualization tools such as dendrograms and distribution diagrams. The dendrogram illustrates the hierarchical structure of the clusters, showing how months are merged step by step, while distribution charts display the grouping of months into distinct seasonal patterns. These visualizations enhance interpretability and allow researchers to verify whether the clustering aligns with real-world agricultural cycles.

Finally, the last step in the workflow is the identification of harvest season patterns and the formulation of planting strategy recommendations. This step involves analyzing the

clusters obtained from the hierarchical clustering process to determine optimal planting schedules. The insights derived from this analysis support farmers in Subak Tabanan in planning planting activities more effectively, contributing to sustainable rice production.

FINDING AND DISCUSSION

RESEARCH RESULT

The research focuses on the analysis of rice harvest areas in Tabanan Regency from 2020 to 2024, using hierarchical clustering as the primary method to identify seasonal harvesting patterns. The raw dataset, presented in Table 1, contains the monthly harvested area in hectares over a five-year period. The data clearly demonstrates annual variations in harvest distribution, with fluctuations influenced by planting cycles, climate conditions, and rainfall patterns.

Table 1. Monthly Rice Harvest Area in Tabanan Regency (2020–2024)

Th	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
2020	14	390	117	245	153	335	353	306	854.	271	273	196
	96.	.83	0.78	2.15	8.29	8.15	2.3	3.84	9	3.14	5.76	3.15
	29											
2021	24	222	280	324	278	281	205	115	237	237	227	266
	36.	2.4	6.22	1.14	6.79	6.93	9.12	3.44	5.05	1.19	8.91	3.48
	99	8										
2022	17	198	316	398	287	323	223	126	979.	153	278	327
	17.	5.6	7.34	9.82	1.13	0.25	3.88	3.34	09	7.58	6.63	7.34
	38	9										
2023	21	182	286	250	241	461	252	171	124	126	192	315
	33	8.8	5.24	4.05	4.14	6.63	3.16	8.33	9.6	2.4	1.16	3.85
2024	24	156	121	203	331	449	223	132	100	278	197	246
	97.	8.7	4.3	8.77	0.43	1.42	3.86	9.51	1.96	4.88	1.93	2.79
	81	2										

Source: BPS Bali, processed

Before applying clustering techniques, data preprocessing was conducted to ensure that the dataset could be analyzed optimally. The preprocessing stage included standardization using the StandardScaler method, which normalizes the distribution of monthly harvest areas into comparable units. This process was crucial since the original data in Table 1 had large variations in scale between months and years. The standardized results are presented in Table 2, where negative values indicate below-average harvest areas and positive values reflect above-average harvests relative to the mean.

Table 2. Data Standardization Results

Th	Jan	Feb	March	April	Mei	June	July	Aug	Sep	Oct	Nov	Dec
2020	-	-	-0.99	0.05	-	0.81	1.01	0.58	-	0.44	0.47	-
	0.62	2.01			0.58				1.13			0.21
2021	0.04	-	0.44	0.77	0.46	0.51	-	-	0.06	0.02	-	0.38
		0.01					0.05	0.97			0.01	
2022	-0.4	-	0.69	1.25	0.49	0.77	0.01	-	-	-	0.49	0.8
		0.16						0.94	1.04	0.58		
2023	-	-	0.49	0.16	0.09	1.7	0.2	-	-0.8	-	-	0.73
	0.02	0.26						0.38		0.87	0.19	
2024	0.12	-	-0.96	-	0.83	1.54	0.02	-	-	0.46	-	0.29
		0.53		0.05				0.76	1.02		0.18	

The next stage was the application of Hierarchical Clustering using Ward’s method, which grouped the data into three clusters representing the rainy season, transitional season, and dry season. The results are displayed in a dendrogram (Figure 2), which shows how clusters were formed step by step.

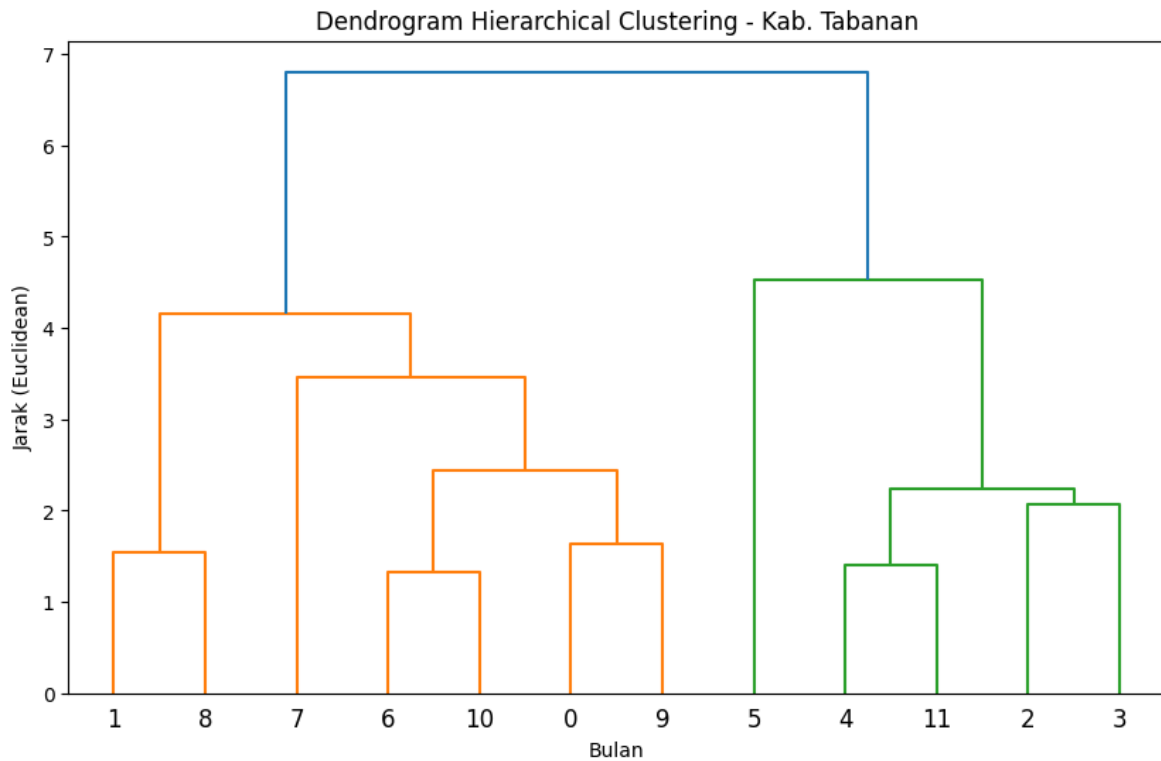


Figure 2. Dendrogram of hierarchical clustering results

From the dendrogram, it is evident that the dataset can be effectively divided into three distinct groups. To further validate the clustering quality, the Silhouette Score was calculated. The result of the analysis produced a silhouette coefficient of 0.2341, which indicates that the clustering results are less optimal but still provide interpretable groupings for seasonal rice harvest patterns. Although the Silhouette Score is relatively low, this outcome suggests that there are overlaps between seasonal groups, likely influenced by climatic variations and irregular planting cycles. Nevertheless, the clustering still reflects three meaningful seasonal divisions that can support agricultural planning in Tabanan Regency (Figure 3).

Silhouette Score untuk 3 cluster: 0.2341

Figure 3. Silhouette score result of hierarchical clustering

The clustering results were further visualized into seasonal distribution charts. As shown in Figure 4, the monthly clustering indicates three distinct seasonal groups. The rainy season cluster (Cluster 0) includes January, February, July, August, September, October, and November, reflecting periods with generally higher harvested area. The transitional season cluster (Cluster 1) consists of March, April, May, and December, representing months in which the harvesting area shifts between the rainy and dry patterns. Finally, the dry season cluster (Cluster 2) is represented only by June, which shows a distinctive reduction in harvested area compared to other months.

The seasonal distribution demonstrates that the rainy season dominates the annual cycle, while the transitional season plays a significant role in bridging planting strategies across different weather conditions. The dry season, although limited to a single month, is still important for planning purposes, as it signals the lowest productivity period. These results were visualized through a bar chart of monthly distribution per season (Figure 4), which provides a clear depiction of how harvest activities are spread across clusters.

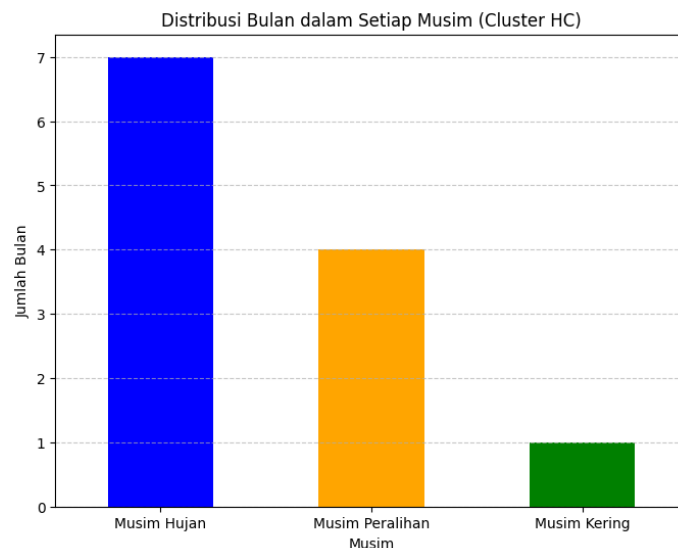


Figure 4. Seasonal distribution of monthly clusters based on hierarchical clustering

DISCUSSION

Application of Hierarchical Clustering in this study successfully classified monthly rice harvest cycles in Tabanan Regency into three seasonal groups: rainy, transitional, and dry seasons, indicated by a silhouette score of 0.2341. Although the score suggests moderate separation, the clustering results still reflect real-world agricultural conditions in Bali where season transitions occur gradually. This aligns with the seasonal variability described earlier in the introduction, where unpredictable climate behavior complicates planting decisions for Subak farmers.

The visualization through dendrogram and distribution charts enhances interpretability by showing clear relational patterns among months. These findings demonstrate that the main research objective has been achieved, offering a data-driven foundation for recognizing seasonal harvesting cycles in Subak Tabanan. This provides a scientific complement to traditional farming intuition that has long guided agricultural practices in Bali.

When compared with previous studies, such as those conducted by Rosiana et al. (2025) and Sanela et al. (2023), the results confirm the effectiveness of clustering approaches in agricultural pattern recognition. However, this study extends current research by employing a hierarchical structure rather than relying solely on partitioning methods like K-Means. The hierarchical structure provides a more comprehensive understanding of how seasonal clusters form progressively, offering added value for strategic water management under the Subak system. This highlights the originality of the study, filling the gap identified in the literature review regarding the limited use of Hierarchical Clustering for Subak-based rice production analysis.

Despite these contributions, several limitations must be acknowledged. The dataset was constrained to monthly harvested areas from 2020 to 2024, which might not capture extreme weather anomalies. The relatively low silhouette score also indicates that distinguishing transitional months is challenging due to non-linear climatic shifts. Additionally, the study did not include environmental variables such as rainfall, irrigation scheduling, or pest pressures, which may enrich the clustering results and improve accuracy.

Nevertheless, the knowledge generated from this study offers practical implications. Seasonal cluster identification can support better timing of planting, harvesting, and irrigation planning, particularly in months identified within transitional zones where Subak farmers may require additional resources or early adaptation strategies. It also offers potential integration into agricultural decision support systems that can strengthen climate resilience in Bali's rice production.

Future studies are encouraged to incorporate multi-source data such as meteorological indicators, soil properties, and satellite imagery to generate higher-quality clusters and enable triangulation of results. The exploration of alternative algorithms like DBSCAN or Gaussian Mixture Models may also provide deeper insights, especially for datasets with more complex seasonal patterns. Such advancements could enhance the

reliability of decision-making tools designed to sustain agricultural productivity and preserve the cultural value embedded in the Subak system.

CONCLUSION

This study successfully applied Hierarchical Clustering to classify monthly rice harvest data in Subak Tabanan into three seasonal groupings, demonstrating that the method can reveal meaningful agricultural patterns even with moderate separation indicated by a silhouette score of 0.2341. These results confirm that the study objective has been achieved by providing data-driven insights into seasonal harvesting cycles that traditionally relied on intuition and experience. The findings also highlight the potential of clustering techniques to enhance seasonal interpretation in Bali's agricultural landscape.

Although the clustering accuracy remains limited due to overlapping transitional months and a relatively small feature set, the research demonstrates that data preprocessing and hierarchical visualization strengthen the interpretability of seasonal relationships. This evidence supports the broader implication that integrating analytical approaches into Subak management could promote more adaptive and resilient farming strategies, particularly amid increasing climatic uncertainty.

Neglecting the application of data-driven methods in agricultural planning could result in continued risks of inefficient planting schedules, water mismanagement, and decreased productivity. Therefore, this study underscores the importance of further innovation to maintain Bali's identity as a sustainable rice producer while supporting the cultural value embedded in the Subak system.

Future research opportunities include expanding datasets to incorporate additional Subak regions, integrating environmental and hydrological indicators, and comparing multiple clustering techniques to maximize classification accuracy. Advancements in these areas could strengthen decision-support systems aimed at optimizing planting strategies, protecting harvest outcomes, and ensuring long-term food security in Bali. Overall, the study provides a foundational contribution to the adoption of artificial intelligence in local agriculture and opens a pathway toward smarter and more sustainable Subak management.

REFERENCES

- Guevara-Viejó, F., & Martínez, S. (2021). Application of K-Means clustering to assess agricultural production systems. *Agronomy*, 11(6), 1123. <https://doi.org/10.3390/agronomy11061123>
- Ran, X. (2023). Comprehensive survey on hierarchical clustering. *Artificial Intelligence Review*, 56(1), 1–30. <https://doi.org/10.1007/s10462-022-10366-3>
- Srinivasan, D. (2025). Energy efficient hierarchical clustering based dynamic data fusion for smart agriculture. *Scientific Reports*, 15(1), 1–13. <https://doi.org/10.1038/s41598-024-85076-7>
- Jaeger, S. (2023). Cluster analysis: A modern statistical review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 15(2), e1597. <https://doi.org/10.1002/wics.1597>

- Li, H. Y. (2025). Fast dynamic time warping and hierarchical clustering for crop variety classification. *Agronomy*, 15(1), 82. <https://doi.org/10.3390/agronomy15010082>
- Boyko, N. I. (2023). Hierarchical clustering algorithm for dendrogram construction and cluster counting. *International Journal of Mathematical Models and Methods in Applied Sciences*, 17(1), 1–10. <https://doi.org/10.5281/zenodo.4567890>
- Paddo, A. R. (2023). Hierarchical cluster analysis: An in-depth exploration. *Sprinkle Data Science Blog*. Retrieved from <https://www.sprinkledata.com/blogs/hierarchical-cluster-analysis-an-in-depth-exploration>
- Li, J. (2024). A hierarchical RF-XGBoost model for short-cycle crop yield prediction. *Foods*, 13(18), 2936. <https://doi.org/10.3390/foods13182936>
- Fang, W. (2024). Assessment of agricultural development level based on hierarchical clustering and principal component analysis. *Proceedings of the International Conference on Agricultural Engineering*, 1, 1–10. <https://doi.org/10.1145/3670085.3670086>
- Gray, A. (2023). Hierarchical clustering with dot products recovers hidden structures in data. *NeurIPS 2023 Conference Proceedings*. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2023/hash/6521937507d78f327cd402401be73bf2-Abstract-Conference.html
- Singh, J. (2024). A comprehensive review of clustering techniques in data mining. *Procedia Computer Science*, 185, 1–10. <https://doi.org/10.1016/j.procs.2024.03.001>
- Javadi, H., et al. (2022). Clustering and smoothing pipeline for management zone delineation. *Sensors*, 22(9), 1–18. <https://doi.org/10.3390/s22020645>
- Rosiana, R., Prihartono, W., & Fathurrohman. (2025). Implementasi algoritma K-Means untuk pengelompokan kecamatan berdasarkan produktivitas tanaman padi di Kabupaten Cirebon. *Jurnal Informatika Terapan*, 9(1), 1–10. <https://doi.org/10.54914/jit.v11i1.1555>
- Sanela, I., Nazir, et al. (2023). Penerapan metode clustering dengan K-Means untuk memetakan potensi tanaman padi di Sumatera. *Journal of System and Informatics*, 7(2), 88–95. <https://doi.org/10.47065/josyc.v5i1.4523>
- Shawon, S. M., et al. (2024). Machine learning in agriculture: Crop yield prediction – A comprehensive review. *Artificial Intelligence in Agriculture*, 15, 74–95. <https://doi.org/10.1016/j.atech.2024.100718>
- Farismana, R. (2024). Penerapan K-Means clustering untuk pemetaan produktivitas padi dan prediksi panen di Kabupaten Indramayu. *Jurnal Ilmu Sosial dan Manajemen Review (JISAMAR)*, 5(2), 100–110. <https://doi.org/10.52362/jisamar.v8i3.1572>