INDONESIAN COUNCIL OF PREMIER STATISTICAL SCIENCE

Indonesian Council of Premier Statistical Science
ISSN: 3030-9956/https://dx.doi.org/10.24014/icopss.v4i2.37870

Volume 4 Issue 2, 81 – 84, August 2025

Analysis of the 2022 Shallot Price Grouping in the Areas of Arengka, Cikpuan, Incense, and Suka Ramai Using the Hierarchical Clustering Method

M. Rizki Akbar[1], Muhammad Angga Piansyah HSB[2], Restu Prasetyo[3]

[1,2]*Prodi Matematika, Faculty of Sciences and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia*

Received: 23 July 2025 Revised: 27 August 2025 Accepted: 30 August 2025 Published: 31 August 2025

Abstract - *This study aims to analyze the price grouping of shallots in four regions, namely Arengka, Cikpuan, Incense, and Suka Ramai in 2022 using the Hierarchical Clustering method. The data used is secondary data on the price of shallots from each region. The analysis was carried out with a multivariate statistical approach to identify price patterns and form clusters based on price similarities between regions. The results of the study show that the price of shallots in the four regions can be grouped into two main clusters. The first cluster consists of areas with lower shallot prices (<66.67 rupiah), while the second cluster consists of areas with higher prices (>66.67 rupiah). The distance between the clusters of 5.22715 shows a significant difference in price characteristics. These findings provide strategic benefits for industry players and policymakers in optimizing distribution and developing marketing strategies that are more targeted in each region. Overall, the Hierarchical Clustering method has proven to be effective in identifying shallot price patterns and can be used as a basis for data-driven decision-making in the agricultural sector.*

Keywords: onion price, hierarchical clustering, cluster analysis, marketing strategy, distribution

1. Introduction

Shallots are one of the horticultural commodities from the food crop sub-sector that have long been destroyed by farmers. Shallots have many uses, especially in the household consumption sector, where shallots can be used as a cooking spice to add flavor to dishes, as complementary ingredients for food, and medicines for certain diseases[1]. Its very strategic role makes shallots classified as one of the basic needs of the Indonesian people. Shallots are also a source of income and employment opportunities that contribute quite high to the economic development of the region[2].

Its role is very strategic to make shallots needed by the people of Indonesia. Shallot consumption in Indonesia continues to increase every year in line with the increase in the population and the improvement of people's living standards and welfare[3]. However, the increase in the amount of demand for shallots is often not offset by the amount of production, considering that shallots are a seasonal, perishable commodity (perishable and not durable). The main obstacles in shallot production include high input costs, pests and diseases, inadequate storage facilities, and limited access to superior seeds. Shallot production also depends on natural conditions such as weather, pests, diseases, air temperature, drought, floods, and natural disasters[4]

2. Research Methods

2.1. *Multivariate Analysis*

Multivariate analysis is a part of statistics that is used to examine data involving more than one variable at the same time. The main purpose of this analysis is to understand the relationship between variables, reduce the dimensions of the data, group the data, or estimate the value of one variable based on other variables[5].

In a study, multivariate analysis is usually applied when the data obtained involves several variables that are interrelated and cannot be examined individually. For example, in socioeconomic research, researchers can observe a person's income, education level, and employment status simultaneously.

- Offers a deeper understanding of complex phenomena.

- Produce a more optimal model by considering the relationships between variables.

Facilitate more informed data-driven decisions

2.2. Definitoin of Cluster Analysis

Cluster analysis is a statistical technique used to group objects based on certain similarities, so that objects in one cluster are more similar to each other than objects in another cluster[3]. The goal is to find hidden structures or patterns in unlabeled data. Cluster analysis is a multivariate statistical technique used to group several objects into several clusters based on similarities or similarities[6].

The goal of this method is to find natural patterns in the data without utilizing dependent variables. Objects with similar characteristics will be grouped into a single cluster, while objects with different characteristics will be separated into another cluster[7]. This approach is exploratory in nature and is commonly applied in various sectors such as marketing, biology, education, and social sciences to identify patterns or relationships hidden in data

2.3 Hierarchical Clustering

One of the most commonly used methods in cluster analysis is hierarchical clustering. This method arranges the data in a stratified structure visualized through a dendogram, which is a tree diagram that illustrates the process of cluster merger[8]. There are two main approaches in hierarchical clustering, namely agglomerative and divisive. The agglomerative approach starts grouping with each object as an individual cluster, then gradually combines based on the degree of similarity, while the divisive approach starts with one large cluster, which is then divided into smaller clusters. The measure of similarity between objects is generally calculated by utilizing Euclidean, Manhattan, or other distance metrics[9].

The process of combining clusters in an agglomerative approach is also influenced by linking methods such as single linkage, complete linkage, average linkage, or the Ward method. The advantage of this method is that it is not necessary to set the number of clusters in advance, but the disadvantage is that it is prone to outliers and less effective when applied to large datasets[10]. The grouping process is carried out in stages, with two main approaches:

- Agglomerative (bottom-up): Each object starts as a separate cluster, then is grouped based on the closest similarity.
- Divisive (top-down): All objects start in a single cluster, then are gradually separated.

Stages of Cluster Analysis

The cluster analysis process generally involves:

1) Data Collection: Preparing the data to be analyzed.
2) Data Preprocessing: Normalizing or standardizing so that variables are at a similar scale.
3) Clustering Method Selection: Choose the right method, e.g. hierarchical or non-hierarchical.
4) Determination of the Number of Clusters: Determine the optimal number of clusters using methods such as the elbow method or silhouette score.
5) Implementation of Clustering: Perform data grouping with a selected algorithm.
6) Evaluation of Results: Evaluate the results of clustering to ensure that the clusters formed are significant and fit for purpose.
7) Interpretation and Reporting: Interpreting the results and compiling a research report.

K-Means is a non-hierarchical cluster analysis method used to group data into clusters based on the similarity of characteristics between objects. The main goal of K-Means is to minimize variation within clusters and maximize variation between clusters, so that each object in one cluster is more similar to each other compared to the objects in other clusters.

2.4 Research Methods

2.4.1 Types of Research

This research is included in the descriptive quantitative category with a multivariate statistical approach, which specifically uses cluster analysis. The purpose of this study is to group the selling prices of shallots in each region.

2.4.2 Data Sources and Types

The data analyzed in this study are secondary data obtained from the Riau Provincial Food and Crop Service. The data includes variables related to the selling price of shallots, including:

- Arengka District(X1).
- Cikpuan District(X2).
- Incense District (X3).
- Suka Ramai District (X4).

3 Analysis

3.1 Research Results

3.1.1 Descriptive Data

Data from this journal was obtained from the Riau Provincial Food and Crop Service

Table 1: Research Variables

Market Average Price/Kg	Moon					
	January	**February**	**March**	**April**	**May**	**June**
Arengka	28	34,1225	33,375	37,625	45,75	55
Miss	27,5	29	29,375	33,25	43,25	45,125
Incense	28,5	34,5	31,875	38,5	46,25	51,875
Likes	25,5	28	28,25	32,625	39,375	46,25
Market Average Price/Kg	Moon					
	July	**August**	**September**	**October**	**November**	**December**
Arengka	49,625	34,625	32,125	33,625	32,125	30
Miss	46,5	34,625	30	30,25	28	29
Incense	49,375	34,125	32,25	32,5	30,625	29,25
Likes	49,25	31,75	30,125	31	29	28,375

This table presents data on the characteristics of the average price of shallots (Thousand rupiah/kg) in the markets of the Arengka, Cikpuan, Incense, and Suka Ramai areas. Each row represents one region determined based on the parameters of the average price of shallots (ten thousand rupiah/kg) in 2022 (January – December).

The data that has been obtained will be processed using MINITAB to form a klister based on the selling price of shallots using the hierarchical method

3.1.2 Application of the Clustering Hierarchy Method

Table 2. Final Score

	Number of observations	**Within-cluster sum of squares**	**Average distance from centroid**	**Maximum distance from centroid**
Cluster1	2	1,89098	0,97236	0,97236
Cluster2	2	6,78590	1,84200	1,84200

The table shows the final result of the application of the Hierarchical Clustering using Minitab software, which Be shown in the form of a "Final Partition" table. This table includes two clusters, each of which has the number of observations, the total square of the distance within the cluster sum of squares, the average distance from centroid, and the maximum distance from centroid. It can be seen that Clusters 1 and 2 have the same number of observations (2) but the value within the cluster sum of squares and the largest average and maximum distances are in cluster 2, this indicates that the distribution of data in cluster two is the most variable. In contrast, Cluster 1 has the smallest distance and square values of distance, indicating that the data in this cluster is highly centered around its centroid. These results provide an overview of the extent of homogeneity of each cluster after data grouping.

Table 3. Distances Between Cluster Centroids

	Cluster1	**Cluster2**
Cluster1	0,00000	5,22715
Cluster2	5,22715	0,00000

The table shows the table of distances between centers (centroids) of the cluster box from the cluster analysis, entitled "Distances Between Cluster Centroids". Each value in this table represents the Euclidean distance between the two centroids of the cluster. It can be seen that the distance between Cluster 1 and Cluster 2 is 5.22715. This greater distance signifies a more significant difference between clusters. This information is important to assess how well the separation between the clusters has been formed in the analysis Based on the results of *the Hierarchical Clustering* analysis of shallot price data, the price of shallots can be grouped into two different clusters. Each cluster has certain characteristics that set it apart from the others. For example, Cluster 1 consists of the price of onions <66.67 (in thousand rupiahs). This shows that the price of shallots in this cluster is low. On the other hand, Cluster 2 consists of the price of onions >66.67 (in thousands of rupiah). This shows that the price of shallots in this cluster is higher than in cluster 1.

Overall, the results of this grouping provide useful information for the industry to understand the differences in the price characteristics of shallots in each region. By knowing the clusters of each shallot price in each region, the distribution process can be regulated more optimally, and marketing strategies can be adjusted to the needs and specifications of each product group.

4 Conclusion

Based on the results of the research conducted on the analysis of the price grouping of shallots in 2022 in the Arengka, Cikpuan, Incense, and Suka Ramai areas using the Hierarchical Clustering method, it can be concluded that the price of shallots in these areas can be grouped into two main clusters. Cluster 1 consists of a lower price of shallots (<66.67 rupiah), while Cluster 2 consists of a higher price of shallots (>66.67 rupiah). The distance between the clusters of 5.22715 shows a significant difference in price characteristics between the two clusters.

This grouping provides important information for industry players and policymakers to understand the price differences in each region. By knowing the price clusters in each region, the distribution process and marketing strategy can be adjusted more optimally and adjusted to the needs and specifications of each region. Overall, the Hierarchical Clustering method has proven to be effective in identifying shallot price patterns, so it can be the basis for planning marketing strategies and more efficient distribution management in the future.

Bibliography

[1] R. A. Rahman, F. M. Afendi, W. Nugraheni, K. Sadik, and A. Rizki, "Grouping and Forecasting of Time Series in Provincial Shallot Production in Indonesia," *National Seminar on Official Statistics*, vol. 2021, no. 1, pp. 457–464, 2021

[2] Stato, "Analysis of Factors Affecting Shallot Price Fluctuations", no. 02, pp. 10-12, 2007. 2

[3] P. S. Agribisnis, F. Sain, D. A. N. Technology, and U. Open, "Independent Research Report on Volatility, Transmission and Integration of Shallot Market Prices in Indonesia Researcher: Argadatta Sigit OPEN UNIVERSITY," 2024

[4] Grema and Gashua, "*The Effect of Shallot Prices on Shallot Production in Central Java*", University of Mataram: Mataram, 2014. 4

[5] I. Nur and L. Fitriana, "Provincial Grouping in Indonesia Based on Healthy Family Indicators Using Hierarchical and Non-Hierarchical Cluster Methods," *J. Paradig. J. Multidisciplinary Mhs. Post-Graduate. Indonesian*, vol. 2, no. 1, pp. 27–36, 2021. 5

[6] A. Nugraha, M. H. Asnawi, P. F. Rahmah, and T. Purwandari, "Analysis of Hierarchy Clusters to Group Provinces in Indonesia based on People's Welfare Indicators," *E-Proceedings of Semin. Nas. Stat. Dep. Stat. FMIPA Univ. Padjadjaran*, vol. 10, no. August, pp. 1–12, 2021, doi: 10.1234/pns.v10i.84. 6

[7] Y. Lasena and Y. Malago, "Clustering of Regional Leading Commodities in Gorontalo Province Using the K-Means Algorithm," *Jambura Journal of Electrical and Electronics Engineering*, vol. 2, no. 1, pp. 14–18, 2020

[8] E. Suprihartiningsih and S. J. A. Samuda, "Hierarchy Cluster Method on Trade Margin Data and Transportation of Strategic Commodities in Indonesia in 2021," *Ecoplan*, vol. 6, no. 1, pp. 8–20, 2023

[9] D. Abid, R. W. Adikusuma, A. M. AL Fikri, and R. K. Hapsari, "Application of the K-Means Clustering Method for Commodity Sales Analysis of Indonesian Farmer Shops," *KERNEL: Journal of Innovation Research in the Field of Informatics and Informatics Education*, vol. 3, no. 2, pp. 25–30, 2023

[10] D. Mulyaningrum, M. Nusrang, and Sudarmin, "Cluster Analysis of the Hierarchical Clustering Method Approach to Economic Growth in South Sulawesi Province," *Makassar State University*, pp. 1–9, 2018