Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

Analysis of the Spatial Distribution of Rice Prices Acrossprovinces in Indonesia Using the K-Means Algorithm

Parini[*] , Yessica Siagian , Guntur Maha Putra , Julpan Hartono Suria Manja Manurung, Nuri Apriani, Putri Wulandari[6]

Faculty of Computer Science, Information Systems Program, Royal University, Kisaran, Indonesia
[4] Faculty of Economics and Law, Law Program, Royal University, Kisaran, Indonesia
Email: [1*] parini.royal@gmail.com, [2] yessica.cyg123@gmail.com, [3] igoenputra@gmail.com, [4] julpanhartonomanurung@gmail.com, [5] nuriapriani@gmail.com , [6] putriwulandari@gmail.com

Abstract—Differences in rice prices across provinces in Indonesia reflect variations in distribution and market conditions that can be analyzed using a data clustering approach. This study aims to map the characteristics of interprovincial rice prices using the K-Means Clustering algorithm. The dataset consists of prices for Premium Rice, Medium Rice, Non-SPHP Medium Rice, and SPHP Rice obtained from official government data. The research stages included data preprocessing, transformation, normalization using StandardScaler, determining the number of clusters using the Elbow Method and Silhouette Score, and visualizing the results using Principal Component Analysis (PCA). The Elbow Method results indicated an efficient point at K = 5, while the highest Silhouette Score of 0.5658 was obtained at K = 7; thus, seven clusters were selected as the best model. PCA visualization reveals that provinces in Indonesia exhibit rice price characteristics that form seven groups with varying degrees of similarity. The research results indicate that the K-Means algorithm is capable of effectively identifying rice price clustering patterns and producing a map that can support the analysis of inter-regional price characteristics as a basis for formulating food distribution policies and stabilizing rice prices in Indonesia.

Keywords: K-Means Clustering, rice prices, clustering, PCA, spatial distribution, Indonesian provinces.

Abstract—Variations in rice prices across Indonesian provinces reflect differences in market conditions and distribution patterns that can be analyzed through a data clustering approach. This study aims to map the characteristics of provincial rice prices in Indonesia using the K-Means clustering algorithm. The dataset consists of the prices of Premium Rice, Medium Rice, Medium Non-SPHP Rice, and SPHP Rice, obtained from official government sources. The research methodology includes data preprocessing, transformation, and normalization using StandardScaler; determination of the optimal number of clusters using the Elbow Method and Silhouette Score; and visualization via Principal Component Analysis (PCA). The Elbow Method indicated an optimal clustering point at K = 5, while the highest Silhouette Score of 0.5658 was achieved at K = 7, making seven clusters the optimal model. The PCA visualization demonstrates that Indonesian provinces can be grouped into seven distinct clusters based on similarities in rice price characteristics. The findings indicate that the K-Means algorithm effectively identifies provincial rice price patterns and provides a spatial clustering model that can support evidence-based decision-making for food distribution planning and rice price stabilization policies across Indonesia.

Keywords: K-Means clustering, rice prices, spatial clustering, principal component analysis, food distribution, Indonesian provinces.

1. INTRODUCTION

Rice is a strategic food commodity that plays a vital role in maintaining national food security, as it serves as the staple food for the majority of the Indonesian population. Rice price stability is a key indicator in preserving public purchasing power and controlling food inflation. However, Indonesia's geographical condition as an archipelago leads to variations in rice prices across provinces, influenced by factors such as production, distribution, logistics costs, and demand levels. These variations indicate that rice price characteristics are not homogeneous across regions, necessitating an analytical approach capable of identifying patterns of similarity and differences among provinces as a basis for formulating more effective food distribution policies.[1]

Advances in data mining technology offer opportunities to process large volumes of food price data into more meaningful information. One widely used technique is clustering, an unsupervised learning method that groups objects based on the similarity of their characteristics. The K-Means algorithm is a popular clustering method because it features a simple, efficient computational process and is capable of producing data clusters with a high degree of homogeneity. To achieve optimal clustering results, the number of clusters is typically determined using the Elbow Method and validated with the Silhouette Score, while Principal Component Analysis (PCA) is used to visualize the clustering results in a low-dimensional space, making them easier to interpret.[2], [3], [4], [5]

Several previous studies have shown that the K-Means algorithm has been widely applied in the food sector. Sinaga et al. clustered 38 provinces in Indonesia based on indicators of consumption, production, rice prices, and population size to support food security analysis, and demonstrated that the clustering approach can provide useful information for formulating food distribution policies. Another study by Christopher et al. combined the K-Means algorithm with OR-Tools to optimize rice distribution at Perum BULOG , resulting in more efficient distribution routes and supporting the stabilization of the national food supply. Additionally, Sipayung and Hasugian integrated PCA and K-Means for the

Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

segmentation of strategic food commodities and demonstrated that the combination of these two methods can enhance the interpretation of clustering results in support of national food policy.[6], [7], [8], [9]
Although various studies have applied the K-Means algorithm to the food sector, most still focus on aspects of production, consumption, sales, or distribution optimization. Research specifically mapping interprovincial rice price characteristics using price data as the basis for spatial clustering remains relatively limited. However, information on price similarity patterns across regions can provide insights into market characteristics and distribution that are useful as input for formulating policies on price stabilization and equitable food distribution.[10], [11], [12], [13]
Based on these issues, this study aims to analyze and map interprovincial rice price characteristics in Indonesia using the K-Means Clustering algorithm. The dataset used includes prices for Premium Rice, Medium Rice, Non-SPHP Medium Rice, and SPHP Rice across all provinces in Indonesia. The research stages include data preprocessing, normalization using StandardScaler, determining the number of clusters using the Elbow Method and Silhouette Score, and visualizing the results with Principal Component Analysis (PCA). The research results are expected to provide information on the patterns of interprovincial rice price clustering as a basis for decision-making in food distribution planning, price stability evaluation, and the formulation of more targeted food policies.

2. RESEARCH METHODOLOGY

2.1 Research Flowchart

The research methodology was systematically designed to cluster Indonesian provinces based on the price characteristics of four types of rice using the K-Means algorithm. The research stages began with data collection, followed by data cleaning, data transformation and aggregation, normalization, determination of the optimal number of clusters, implementation of the K-Means algorithm, evaluation of the clustering results, visualization of the results, and finally analysis and drawing of conclusions. This research workflow was designed so that the clustering process would produce clusters with a high degree of homogeneity within each group and clear heterogeneity between groups. The research methodology flowchart is shown in Figure 1.

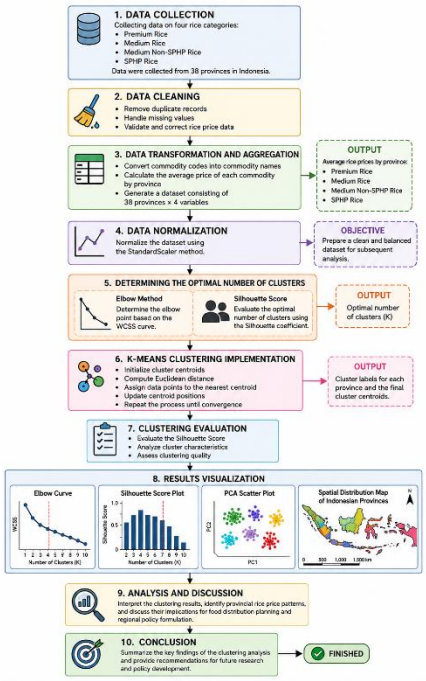

Figure 1. Research Flowchart

Based on Figure 1, the study began with the collection of price data for four types of rice—Premium Rice, Medium Rice, Non-SPHP Medium Rice, and SPHP Rice—across all provinces in Indonesia. The obtained data then underwent a data cleaning process to remove duplicate entries, handle missing values, and validate the price format, ensuring all data was ready for analysis.
The next stage involved data transformation and aggregation, specifically converting commodity codes into commodity names and calculating the average price for each type of rice in each province. The result of this process formed a dataset with one row for each province and four rice price variables as analysis attributes. Next, data normalization was performed

Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

using the StandardScaler method to ensure all variables had the same scale, thereby preventing any single attribute from dominating the clustering process.
After the normalization process is complete, the optimal number of clusters is determined using the Elbow Method and the Silhouette Score. The Elbow Method is used to determine the number of clusters based on the Within-Cluster Sum of Squares (WCSS) value, while the Silhouette Score is used to evaluate the quality of separation between clusters. The number of clusters that yields the best evaluation results is then used as a parameter in the implementation of the K-Means algorithm.
During the implementation of the K-Means algorithm, the system initializes the centroids, calculates the Euclidean distance between each data point and the centroids, assigns data points to the nearest centroid, updates the centroid positions, and repeats this process until the centroids reach a state of convergence. The result of this process is a set of provincial clusters along with centroid values that represent the characteristics of each cluster.
Next, the clustering results are evaluated through an analysis of the Silhouette Score and the characteristics of each cluster to ensure the quality of the resulting groupings. The clustering results are then visualized in the form of an Elbow Method graph, a Silhouette Score graph, a scatter plot using Principal Component Analysis (PCA), and a map showing the distribution of provincial clusters in Indonesia. These visualizations facilitate the interpretation of the spatial distribution patterns of rice prices.
The final stage of the study is the analysis and discussion, which involves interpreting the characteristics of each cluster based on the obtained clustering results and relating them to the distribution of rice prices across provinces. Based on these analysis results, conclusions were drawn that include the study's main findings along with recommendations that can serve as input for the government and stakeholders in supporting food distribution policies and rice price control in Indonesia.

2.3 K-Means Algorithm

The K-Means algorithm is one of the most widely used *unsupervised learning* methods for clustering data based on the similarity of characteristics among objects. The primary goal of K-Means is to partition a dataset into K clusters such that objects within the same cluster share a high degree of similarity, while objects in different clusters have a low degree of similarity. This algorithm works iteratively by minimizing the sum of the squared distances between each data point and the cluster center (*centroid*), thereby producing more homogeneous groups. K-Means is known for its simple and efficient implementation, as well as its ability to handle large datasets [14], [15], [16].
In this study, the K-Means algorithm was used to cluster provinces in Indonesia based on the price characteristics of four types of rice: Premium Rice, Medium Rice, Non-SPHP Medium Rice, and SPHP Rice. The clustering results are expected to reveal patterns in the spatial distribution of food prices, thereby serving as a basis for identifying regional characteristics with relatively low, moderate, or high price conditions.[9], [17], [18], [19], [20]
The objective of the K-Means algorithm is to minimize the sum of squared distances within each cluster (*Within-Cluster Sum of Squares* or WCSS).

$$J = \sum_{j=1}^{k} \sum_{xi_i \in C_j} ||x_i - c_j||^2$$

Notes:

J = K-Means objective function
K = number of clusters
Cj = the jth cluster
Cj = centroid of the jth cluster

The smaller the value of the objective function, the better the quality of the resulting clustering, because the members within each cluster are more homogeneous.

3. RESULTS AND DISCUSSION

3.1 Data Preprocessing Results

The preprocessing stage is the initial step taken to improve data quality before applying the K-Means algorithm. This process aims to ensure that the data used meets the requirements for analysis, so that the clustering results obtained are more accurate and scientifically valid. The preprocessing stages in this study include data structure inspection, data cleaning, attribute transformation, data aggregation, and data normalization.

Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

3.1.1 Description of the Initial Dataset

The dataset used consists of interprovincial rice price data in Indonesia, comprising four types of commodities: Premium Rice, Medium Rice, Non-SPHP Medium Rice, and SPHP Rice. The data is organized by province and year of observation, so that each province has multiple data points corresponding to the commodity type and recording period. Overall, the initial dataset consists of 5,146 observations with five attributes: Province Code, Province Name, Commodity, Year, and Price. The Price attribute is the primary variable used in the clustering process, while the other attributes serve to identify regions and commodity categories, as shown in Table 1.

Table 1 Characteristics of the Initial Dataset

No	Characteristic	Value
1	Number of Observations	**5,146**
2	Number of Attributes	**5**
3	Attributes	Province Code, Province Name, Commodity, Year, Price
4	Number of Commodities	**4**
5	Commodity Type	Premium Rice, Medium Rice, Non-SPHP Medium Rice, SPHP Rice
6	Unit Price	Rupiah/kg

3.1.2 Data Cleaning Results

Before the clustering process was performed, the dataset underwent a data cleaning stage to eliminate data inconsistencies that could affect the analysis results. This stage included checking for missing values, validating data types, removing non-numeric characters from the price attribute, and transforming commodity codes into commodity names.

Additionally, data aggregation was performed by calculating the average price of each commodity in each province. This process produced one observation for each province, which was subsequently used as input for the K-Means algorithm. After the aggregation process was complete, the data was normalized using the StandardScaler method to ensure that all variables had a uniform scale and did not introduce bias during the clustering process.

3.2 Clustering Results Using the K-Means Algorithm

After the data *preprocessing* and normalization stages were completed, the next step was to apply the K-Means algorithm to cluster Indonesian provinces based on the average prices of four types of rice: Premium Rice, Medium Rice, Non-SPHP Medium Rice, and SPHP Rice. Before the clustering process began, the optimal number of clusters was determined using the Elbow Method and the Silhouette Score.

3.2.1 Creation of the Clustering Dataset

The data, which was originally in transaction format, was converted into a pivot table so that each province is represented as a single observation with four rice price variables. An example of the transformed data is shown in Table 2

Table 2 Results of Data Transformation

Province	Medium Rice	Medium Non-SPHP	Premium Rice	SPHP
Aceh	14,046.27	14,575.00	15,362.91	12,876.27
Bali	14,089.64	14,246.00	15,818.36	11,999.45
Banten	12,944.73	12,897.00	14,852.73	12,188.00
Bengkulu	13,543.27	13,699.00	15,589.09	12,393.09
…..	…..	……	…..	…..
Special Region of Yogyakarta	13,052.73	13,266.00	14,393.09	12,446.82

Table 3 shows that each province has different price characteristics for the four types of rice. These differences form the basis for the clustering process using the K-Means algorithm.

3.2.2 Determining the Number of Clusters

The number of clusters was determined using the Silhouette Score by testing several K values ranging from 2 to 8.

Table 3. Silhouette Score Test Results

Number of Clusters	Silhouette Score
2	0.4848
3	0.4976
4	0.5369
5	0.5631
6	0.5564

Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

7	0.5658
8	0.5636

Based on Table 3, the Silhouette Score increases as the number of clusters increases. The highest value was obtained at K = 7, which was 0.5658, indicating that clustering with seven clusters provides the best separation quality compared to other numbers of clusters. In general, a Silhouette Score above 0.5 indicates that the resulting cluster structure is sufficiently good because it exhibits a high degree of homogeneity within clusters and clear heterogeneity between clusters.

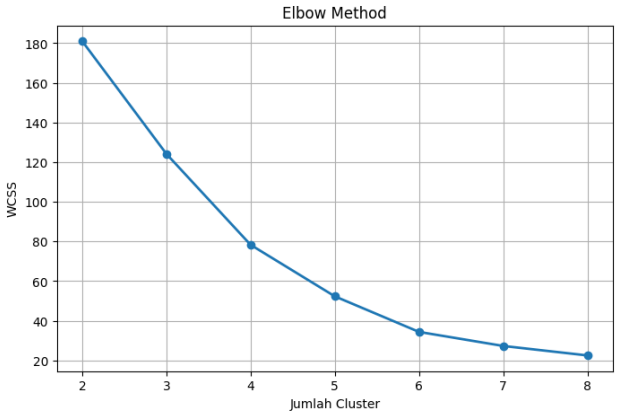

Figure 2. Elbow Method

Based on Figure 2, it can be seen that the Within-Cluster Sum of Squares (WCSS) value decreases as the number of clusters used increases. When the number of clusters is K = 2, the WCSS value is still quite high, at around 181, indicating that the variation within clusters is still substantial. When the number of clusters was increased to K = 3, K = 4, and K = 5, the WCSS values decreased significantly to approximately 124, 78, and 53. This sharp decrease indicates that increasing the number of clusters improves the homogeneity of the data within each cluster.

Furthermore, for the number of clusters K = 6, K = 7, and K = 8, the decrease in the WCSS value begins to slow down, with values of approximately 35, 27, and 23, respectively. This indicates that adding clusters beyond K = 5 no longer results in a significant reduction in data variation. Thus, based on the pattern in the graph, the *elbow point* begins to form at K = 5, so this number of clusters can be considered the optimal number based on the Elbow method.

However, this study does not rely solely on the Elbow Method to determine the optimal number of clusters. Evaluation results using the Silhouette Score show that the highest value is obtained at K = 7 with a value of 0.5658, indicating that forming seven clusters provides better data separation quality compared to other numbers of clusters. Therefore, although the Elbow Method indicates K = 5 as the efficient point, this study uses K = 7 as the final number of clusters because it provides more optimal clustering quality based on the Silhouette Score evaluation.

These results demonstrate that the combined use of the Elbow Method and the Silhouette Score provides a stronger basis for determining the optimal number of clusters. The Elbow Method is used to identify the efficient number of clusters in terms of reducing data variation, while the Silhouette Score is used to evaluate the quality of the resulting cluster structure. Thus, the selection of K = 7 is considered more representative for grouping Indonesian provinces based on rice price characteristics, so that the clustering results can provide more accurate information in the analysis of the spatial distribution of food commodity supply gaps.

Thus, based on the evaluation results using the Silhouette Score, the recommended number of clusters is seven (K = 7).

3.2.3 Clustering Results

After the normalization process was completed, the data were clustered using the K-Means algorithm based on four variables: Medium Rice, Non-SPHP Medium Rice, Premium Rice, and SPHP Rice. The dataset used in the clustering process was derived from a transformed dataset (pivot table), so that each observation represents the rice price characteristics of its respective province.

Table 5. Clustering Results

No	Province Name	Medium Rice	Non-SPHP Medium Rice	Premium Rice	Cluster
1	Aceh	14,046.27	13,017.83	15,362.91	C1
2	Bali	14,089.64	12,186.67	15,818.36	C1
3	Banten	12,944.73	12,247.08	14,852.73	C7
4	Bengkulu	13,543.27	12,501.92	15,589.09	C7
5	Special Region of Yogyakarta	13,052.73	12,515.08	14,393.09	C7
33	Central Sulawesi	...	...	...	...
34	Southeast Sulawesi	11,535.36	12,301.00	13,064.92	C2

Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

35	North Sulawesi	12,160.22	12,429.22	13,346.92	C2
36	West Sumatra	13,400.76	12,947.67	15,176.60	C7
37	South Sumatra	11,333.56	12,492.67	12,755.24	C2
38	North Sumatra	12,338.90	12,698.00	13,623.92	C2

3.2.4 Visualization of Provincial Clustering Results

After the optimal number of clusters was determined using the K-Means algorithm, the clustering results were visualized using Principal Component Analysis (PCA). PCA was used to reduce the four research variables—the prices of Medium Rice, Non-SPHP Medium Rice, Premium Rice, and SPHP Rice—into two principal components (*Principal Component 1* and *Principal Component 2*). This visualization aims to facilitate the interpretation of clustering patterns among provinces.

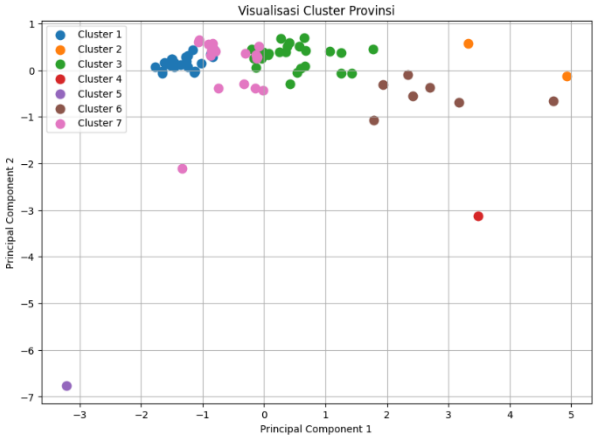

Figure 3: Visualization of Provincial Clustering Results Using K-Means (K = 7)
Source: Data Processing Results Using Python (2026).

Based on Figure 3, it can be seen that the K-Means algorithm successfully grouped the data into seven clusters, each displayed in a different color. Each point on the graph represents a provincial observation, while the distance between points indicates the level of similarity in rice price characteristics. The closer the points are to each other, the higher the level of price similarity between provinces.

Cluster 1 (blue) has a relatively large number of members and forms a tightly clustered group on the left side of the graph. This indicates that the provinces in this cluster have relatively homogeneous price characteristics with low variation.

Cluster 2 (orange) consists of only a few provinces located far to the right of the graph. Their isolated position from the other clusters indicates that the provinces in this group have price characteristics that are quite different from those of the majority of other provinces. These differences may be influenced by high average prices for certain commodities or differing food distribution conditions.

Cluster 3 (green) forms a group with a fairly large number of members spread across the middle of the graph. This distribution indicates greater price variation compared to Cluster 1, yet the provinces still share a sufficiently high degree of similarity to be grouped together.

Cluster 4 (red) consists of only one province, located quite far from the other groups. This position indicates that the province has price characteristics that are so distinct that it forms its own cluster. In cluster analysis, this condition indicates the presence of an outlier—a region with a unique price pattern.

Cluster 5 (purple) also consists of only one province, located at the bottom-left corner of the graph. Its significant distance from the other clusters indicates that this province has price characteristics that differ the most from all other provinces. These differences may be caused by geographical conditions, high logistics costs, supply constraints, or local economic factors.

Cluster 6 (brown) consists of several provinces located on the right side of the graph with a relatively close distribution pattern. This indicates that the provinces in this cluster share nearly identical price characteristics and differ from the other groups of provinces.

Meanwhile, Cluster 7 (pink) is situated between Cluster 1 and Cluster 3. This position indicates that the provinces in this cluster have price characteristics at an intermediate level, making them a transitional group between provinces with relatively low prices and those with relatively high prices.

Overall, the visualization shows that most provinces were successfully grouped into clusters with fairly distinct distribution patterns. Although there are several clusters consisting of only one province, this indicates the presence of regions with price characteristics that are significantly different from those of other regions. This information suggests that the distribution of rice prices in Indonesia is not homogeneous but rather forms several groups with distinct characteristics.

Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

The results of this visualization reinforce the findings of the evaluation using the Silhouette Score, which show that using seven clusters (K = 7) produces a good separation of the data. Therefore, the clustering results can serve as a basis for identifying regions with similar price characteristics, thereby supporting the formulation of policies for food distribution, price stabilization, and the equitable distribution of rice supplies in Indonesia.

3.3 Discussion

Based on Table 5, the highest Silhouette Score was obtained at K = 7 with a value of 0.5658. This value indicates that forming seven clusters yields the best level of homogeneity within clusters and heterogeneity between clusters compared to other numbers of clusters.

The evaluation results reveal a discrepancy between the Elbow Method and the Silhouette Score. The Elbow Method indicates that the optimal number of clusters is K = 5, whereas the Silhouette Score shows that K = 7 yields better cluster separation quality.

In this study, the number of clusters used should be determined based on the Silhouette Score, namely K = 7, because this method directly evaluates the quality of the clustering results. The Silhouette Score value of 0.5658 also indicates that the resulting cluster structure is sufficient for grouping provinces based on rice price characteristics.

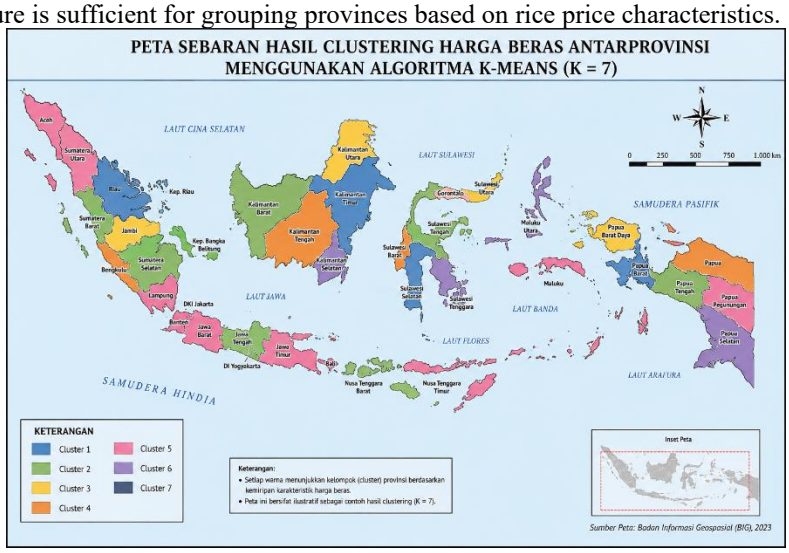

Figure 4. Distribution of Rice Price Clusters

4. CONCLUSION

This study successfully applied the K-Means Clustering algorithm to map the characteristics of rice prices across provinces in Indonesia as an indicator related to the spatial distribution of food supply gaps. Before the clustering process was carried out, the data underwent preprocessing stages that included data cleaning, transformation, aggregation, and normalization to ensure the quality of the data used in the analysis. The number of clusters was determined using the Elbow Method and the Silhouette Score. The results of the Elbow Method indicated that the efficient point occurred at K = 5, while the evaluation using the Silhouette Score yielded the highest value of 0.5658 at K = 7; thus, seven clusters were selected as the best model because they provided better cluster separation. Visualization using Principal Component Analysis (PCA) showed that provinces in Indonesia can be grouped into seven clusters based on the similarity of rice price characteristics, with most clusters exhibiting clear distribution patterns, although there were a few clusters with very few members, reflecting price characteristics that differed from those of the majority of provinces. The results of this study indicate that the K-Means algorithm is effective in identifying patterns of rice price similarity across regions and produces a map that can serve as a basis for decision-making in food distribution planning, price stability control, and policy prioritization in regions with price characteristics that differ from those of other provinces.

Acknowledgements

This research was funded by the Directorate General of Higher Education, Research, and Technology of the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia through the 2026 Early-Career Research Grant Program. The authors express their sincere gratitude for the support provided.

Bulletin of Information Technology (BIT)
Vol 7, No 3, September 2026, Hal. 351 – 358
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.3146
https://journal.fkpt.org/index.php/BIT

REFERENCES

[1] H. V. Christopher, A. A. Purnama, and S. M. M. Harahap, "Application of K-Means Clustering and OR-Tools to Optimize Rice Distribution: A Case Study of Perum Bulog Indonesia," *Applied Information System and Management (AISM)*, vol. 7, no. 2, pp. 57–63, Sep. 2024, doi: 10.15408/AISM.V7I2.40618.

[2] M. Fariz Fadillah Mardianto *et al.*, "Time Series Clustering Analysis for Increases Food Commodity Prices in Indonesia Based on K-Means Method," *Journal of Human, Earth, and Future*, vol. 5, no. 3, pp. 319–329, Sep. 2024, doi: 10.28991/HEF-2024-05-03-02.

[3] A. Karim *et al.*, "Comparison Of Machine Learning Algorithms For Rice Production Prediction," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 10, no. 2, pp. 407–418, Apr. 2026, doi: 10.29207/RESTI.V10I2.7453.

[4] A. Karim *et al.*, "Optimizing Agricultural Commodity Price Forecasts Using an Ensemble Stacking Method Based on Market and Product Characteristics," *Journal Global Technology Computer*, vol. 5, no. 3, pp. 127–135, Aug. 2026, doi: 10.47065/JOGTC.V5I3.11033.

[5] A. Karim and A. P. Juledi, "PCA-Assisted K-Means Clustering of 31 Food Commodity Price Profiles in Indonesia," *International Journal of Informatics and Data Science*, vol. 3, no. 1, pp. 23–33, Decs. 2025, doi: 10.64366/IJIDS.V3I1.543.

[6] R. F. Sinaga, M. A. Prabukusumo, and J. Manurung, "Comparison of k-means clustering with hierarchical agglomerative clustering for the analysis of food security of rice sector in Indonesia," *Journal of Intelligent Decision Support System (IDSS)*, vol. 8, no. 1, pp. 22–33, Mar. 2025, doi: 10.35335/IDSS.V8I1.290.

[7] J. Suryadi, "CLUSTERING KETIDAKCUKUPAN KONSUMSI PANGAN PER KABUPATEN/KOTA MENGGUNAKAN ALGORITMA K-MEANS DAN FUZZY C-MEANS," *Jurnal Komputer dan Informatika*, vol. 19, no. 1, pp. 30–35, Apr. 2024, Accessed: Aug. 06, 2026. [Online]. Available: https://journal.untar.ac.id/index.php/JKI/article/view/34579

[8] 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.

[9] S. Retno, Bustami, and R. K. Dinata, "Enhancing K-Means Clustering Model to Improve Rice Harvest Productivity Areas in Aceh Utara Using Purity," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 13, no. 2, pp. 405–418, Jul. 2024, doi: 10.23887/JANAPATI.V13I2.78254.

[10] S. Pardingotan Sipayung and P. Marto Hasugian, "Integration Of Pca And K-Means Clustering For Staple Food Segmentation In Support Of National Food Policy," *Sinkron : jurnal dan penelitian teknik informatika*, vol. 9, no. 4, pp. 3175–3189, Oct. 2025, doi: 10.33395/SINKRON.V9I4.15343.

[11] M. Veronica, H. Effendi, and A. Oktafian Saleh, "Clustering Tingkat Kedisiplinan Pegawai Pada Pengadilan Tinggi Palembang Menggunakan Algoritma K-Means," in *SEMINAR NASIONAL CORISINDO*, 2023, pp. 261–266.

[12] R. E. Marpaung, "Penerapan Algoritma K-Means Dalam Mengclustering Kualitas Bibit Kelapa Sawit Di PPKS Marihat," *Program Studi Sistem Informasi*, vol. 1, no. 1, pp. 7–15, 2022.

[13] I. Melani, B. Priyatna, F. Nurapriani, and S. S. Hilabi, "Implementasi Metode K-Means Clustering Pada Penilaian Kinerja Karyawan PT Kopetri Citra Abadi," *Jurnal Informasi Interaktif*, vol. 8, no. 1, pp. 24–30, 2023.

[14] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982, doi: 10.1109/TIT.1982.1056489.

[15] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithms: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002, doi: 10.1109/TPAMI.2002.1017616.

[16] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics 2020, Vol. 9, Page 1295*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/ELECTRONICS9081295.

[17] W. Purba, W. Siawin, and . H., "Implementasi Data Mining Untuk Pengelompokkan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering," *Jurnal Sistem Informasi dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019, doi: 10.34012/jusikom.v2i2.429.

[18] U. Surapati and M. Jannah, "Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer Dalam Pembelian Merchandise Kpop," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 875–884, 2024, doi: 10.55338/saintek.v5i3.2739.

[19] G. Karagiannis and P. Ravanos, "On the use of Malmquist productivity indices for intertemporal performance assessment by means of composite indicators," *The Journal of Economic Asymmetries*, vol. 31, p. e00404, Jun. 2025, doi: 10.1016/J.JECA.2025.E00404.

[20] P. Di, P. Dki, and F. Handayanna, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Penerapan Algoritma K-Means Untuk Mengelompokkan Kepadatan," vol. 5, no. 1, pp. 50–55, 2024.