

PCA-Assisted K-Means Clustering of 31 Food Commodity Price Profiles in Indonesia

Abdul Karim^{1,*}, Angga Putra Juledi²

¹Faculty of Science and Technology, Department of Information Technology, Universitas Labuhanbatu, Rantauprapat, Indonesia

²Faculty of Science and Technology, Department of Information Systems, Universitas Labuhanbatu, Rantauprapat, Indonesia

Email: ^{1,*}abdulkarim@ulb.ac.id, ²anggapj19@ulb.ac.id

Corresponding Author Email: abdulkarim@ulb.ac.id

Received: 10/12/2025, Revised: 17/12/2025, Accepted: 30/12/2025, Published: 31/12/2025

Abstract—Food price volatility exhibits heterogeneous characteristics across commodities, making uniform monitoring and stabilization strategies potentially ineffective. This study proposes a multidimensional clustering framework to identify homogeneous price profiles among 31 food commodities in Indonesia using Principal Component Analysis (PCA) and K-Means clustering. Daily commodity prices observed from September 2, 2019, to January 17, 2020, were transformed into statistical price characteristics, including mean, standard deviation, minimum, maximum, coefficient of variation, price changes, and volatility. The features were standardized using Z-score normalization and subsequently transformed using PCA to reduce redundancy while preserving the majority of information. The first ten principal components retained 97.39% of the total variance. Candidate cluster solutions were evaluated using the Elbow Method, Silhouette Score, Davies–Bouldin Index, and Calinski–Harabasz Index, resulting in an optimal three-cluster solution based on the combined validation criteria. The resulting clusters represent distinct multidimensional price profiles characterized by differences in price level, dispersion, price variation, and volatility. The proposed framework contributes a data-driven approach for grouping food commodities according to their joint price characteristics rather than considering price level alone. The resulting profiles can support differentiated monitoring priorities and more targeted food price stabilization policies.

Keywords: Food Commodity Prices; Price Volatility; Principal Component Analysis; K-Means Clustering; Cluster Validation; Food Price Stabilization

1. INTRODUCTION

Food commodity prices are an important determinant of food security, household purchasing power, and market stability. Persistent price fluctuations can increase uncertainty for consumers, producers, distributors, and policymakers, particularly when price movements differ substantially across commodities. Recent evidence from Indonesia indicates that strategic food commodities exhibit heterogeneous levels of price fluctuation and volatility, suggesting that food price stabilization cannot necessarily rely on a uniform intervention strategy [1], [2]. In the Indonesian context, the vast geographical expanse and inter-regional distribution structure mean that price change patterns across commodities may exhibit distinct characteristics. Research on strategic food commodities in Indonesia also indicates differences in the degree of fluctuation across commodities, with some horticultural commodities tending to exhibit higher volatility compared to certain food commodities that are relatively stable [3],[4].

These issues indicate that food price analysis cannot be adequately conducted based solely on average values or individual price changes. An approach is needed that can identify similarities in characteristics among commodities based on multiple price indicators simultaneously. A clustering approach can be used to group objects based on the degree of similarity in their characteristics without requiring prior class labels. One widely used algorithm for this purpose is K-Means, which performs clustering based on the proximity of objects to the cluster center (centroid) and iteratively minimizes the variance within clusters [5]. This approach allows commodities with relatively similar price patterns to be placed in the same cluster, enabling a more systematic analysis of each cluster's characteristics [6], [7].

However, the use of many price characteristic variables can result in a relatively high data dimension and potentially lead to information redundancy among variables. Therefore, Principal Component Analysis (PCA) can be used as a dimension reduction approach prior to the clustering process [8]. PCA transforms a set of initial variables into principal components, which are linear combinations of those variables, and ranks them based on their ability to explain data variance [9], [10]. Thus, PCA can produce a more concise representation of the data while retaining most of the essential information. The use of PCA can also aid in visualizing the data structure, making the clustering patterns among objects easier to interpret [11].

Determining the number of clusters is a critical aspect in applying K-Means because too few clusters can result in an unrepresentative grouping, while too many can increase the complexity of interpretation. The Elbow Method can be used by evaluating the within-cluster sum of squares or inertia for several alternative numbers of clusters. In addition, the Silhouette Score can be used to evaluate the degree of cohesion of objects within clusters and the separation between clusters [12], [13]. These evaluation approaches provide a more objective basis for determining a cluster structure that aligns with the data's characteristics [14], [15], [16].

A number of previous studies have examined the volatility and dynamics of food prices in Indonesia using statistical and econometric approaches. Sari et al. analyzed the volatility characteristics of food commodity prices in Indonesia and demonstrated differing volatility patterns across commodities [17]. Matondang et al. also

demonstrated that strategic food commodities exhibit different levels of fluctuation and volatility, thus requiring management approaches tailored to the characteristics of each specific commodity[18]. More recent research indicates that the volatility of food producer prices may also exhibit inter-regional correlations, meaning that price characteristics are not always isolated to a single region or a single commodity [19]. Furthermore, research on rice price transmission in Indonesia shows that price changes at the market level may be linked to price dynamics at other levels of the supply chain [20].

Although numerous studies on food price volatility and transmission have been conducted, approaches that specifically cluster food commodities based on combinations of price characteristics using PCA and K-Means still have room for development. This study addresses this gap by constructing a representation of price characteristics through feature engineering, standardization, and dimensionality reduction using PCA, followed by clustering using K-Means.

This study aims to identify patterns of similarity in price characteristics and to cluster food commodities into groups with relatively homogeneous characteristics using a combination of PCA and K-Means clustering. The number of clusters was determined using the Elbow Method, while the clustering results were subsequently analyzed based on the feature characteristics of each cluster. The results of this study are expected to contribute to the application of unsupervised learning methods for food price analysis and to provide information that can support monitoring, the identification of commodity characteristics, and the formulation of food price stabilization strategies.

2. RESEARCH METHODOLOGY

2.1 Research Stages

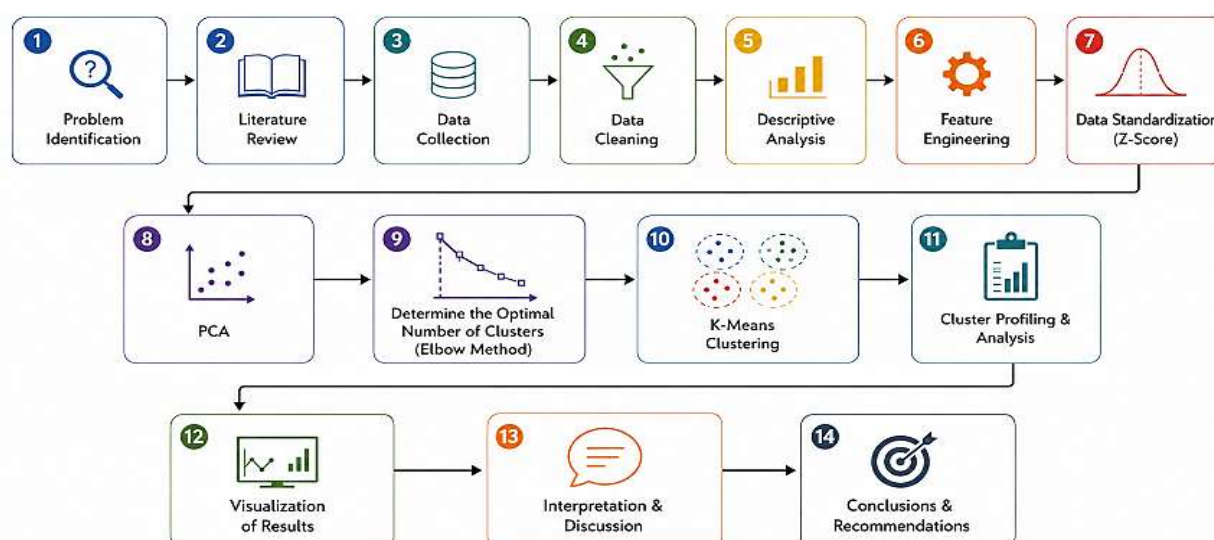


Figure 1. Research Flowchart

Figure 1 illustrates the problem identification stage, which involves examining differences in price levels and fluctuations across various food commodities. These varying characteristics result in price patterns that are not always consistent across commodities, thus requiring a data mining approach to identify groups of commodities with similar price characteristics. Next, data were collected in the form of daily price data for 31 food commodities during the period from September 2, 2019, to January 17, 2020. The data obtained then underwent a data preprocessing stage, which included checking and handling missing values, duplicate data, data types, price formats, and data consistency to produce a clean dataset ready for analysis.

The next stage involved feature engineering and standardization, which entailed creating a set of features representing the price characteristics of each commodity, such as average price, price change, standard deviation, minimum and maximum values, coefficient of variation, and price volatility. The resulting features are then standardized using the Z-Score method to normalize the scale across variables. Next, Principal Component Analysis (PCA) is applied to reduce the data dimensions and help illustrate the structure of relationships among commodities in a simpler manner. Afterward, the optimal number of clusters is determined using the Elbow Method and Silhouette Score to identify the number of groups that yields the best clustering structure.

Based on the optimal number of clusters obtained, the K-Means clustering process was performed by grouping commodities based on the similarity of their price characteristics. The K-Means algorithm operates iteratively by determining centroids, calculating the distance of each object from the centroids, forming groups based on the nearest centroid, and updating the centroids until a stable result is obtained. The final stage involves the analysis and interpretation of the clusters, specifically identifying the characteristics of each group based on feature

values, the degree of variation, and price volatility. The clustering results are then visualized using PCA scatter plots, elbow plots, silhouette plots, and cluster characteristic visualizations. These results are subsequently used to identify commodity groups with relatively stable prices as well as those exhibiting high volatility, serving as the basis for formulating the study's conclusions and recommendations.

2.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimension reduction method used to simplify a set of variables into several principal components while retaining most of the information or variance in the data. In this study, PCA was applied after standardizing the food commodity price features.[21]

$$Z_{ij} = \frac{X_{ij} - X_j}{S_j} \quad (1)$$

Where (Z_{ij}) represents the standardized value of the (i) th object for the (j) th variable, (X_{ij}) denotes the original value of the (i) th object for the (j) th variable, (X_j) represents the mean of the (j) th variable, and (S_j) denotes the standard deviation of the (j) th variable. The standardization process transforms the original observations into values with a mean of zero and a standard deviation of one, thereby eliminating differences in measurement scales among variables. This transformation ensures that each variable contributes comparably to subsequent statistical analyses and prevents variables with larger numerical scales from disproportionately influencing the results.

2.3 Cluster Validation

The optimal number of clusters was determined using multiple internal validation criteria rather than relying exclusively on the Elbow Method. Candidate values of K from 2 to 10 were evaluated using the Silhouette Score, Davies–Bouldin Index, and Calinski–Harabasz Index. The Elbow Method was additionally used to examine the reduction in within-cluster sum of squares (WCSS). The final value of K was selected based on the consistency of the validation results.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

The silhouette coefficient, denoted as $(S(i))$, represents the silhouette value of the (i) -th object or observation and is used to evaluate the quality of clustering. The value $(a(i))$ is the average distance between the (i) -th object and all other objects within the same cluster, representing the degree of cluster cohesion. Meanwhile, $(b(i))$ is the minimum average distance between the (i) -th object and the objects belonging to the nearest neighboring cluster, indicating the degree of separation between clusters. The term $(\max(a(i), b(i)))$ represents the larger value between $(a(i))$ and $(b(i))$ and is used as the denominator to normalize the silhouette coefficient. Thus, the silhouette value provides an overall measure of how well an observation is assigned to its cluster, with higher values indicating better cohesion within the cluster and greater separation from neighboring clusters.

2.4 K-Means Clustering

K-Means Clustering is an unsupervised learning method used to group data based on the similarity of characteristics among objects. In this study, K-Means was applied to cluster food commodities based on price characteristics after undergoing standardization and dimensionality reduction using PCA. The number of clusters was determined based on the results of the Elbow Method, namely $K=3$. Clustering is performed iteratively by calculating the distance of each object from its centroid using Euclidean distance, then assigning the object to the cluster with the shortest distance. The centroid of each cluster is subsequently updated based on the average value of the cluster members until there is no change in membership or convergence is achieved.[22], [23], [24], [25].

Mathematically, the Euclidean distance between object x_i and centroid C_j is defined as:

$$d(x_i, C_j) = \sqrt{\sum_{k=1}^p (x_{ik} - C_{jk})^2} \quad (3)$$

Next, the new centroids are calculated using:

$$C_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i \quad (4)$$

The K-Means process aims to minimize the Within-Cluster Sum of Squares (WCSS):

$$WCSS = \sum_{j=1}^k \sum_{x_i \in C_j} (x_{ik} - C_{jk})^2 \quad (5)$$

The final result of the clustering process yields three commodity groups, namely Cluster 0, Cluster 1, and Cluster 2, which are then analyzed based on price characteristics to identify patterns and differences between clusters.

3. RESULTS AND DISCUSSION

3.1 Data Loading

The data loading stage was performed by importing the food commodity price dataset into the system in CSV format. The loaded data was then examined for its structure, number of attributes, number of observations, and data types to ensure the dataset was successfully read and ready for use in the preprocessing stage. See Table 1.

Table 1. Dataset Loading

No	Date	Rice	Grade I Rice (kg)	Grade I Rice (kg)	Grade I Medium Rice (kg)	Grade II Medium Rice (kg)	Super Grade I Rice (kg)	Super II Quality Rice (kg)	Chicken Meat	Fresh Chicken (kg)	...	Local Granulated Sugar (kg)
0	09/02/2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,450	30,450	...	12,900
1	03/09/2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,300	30,300	...	12,900
2	04/09/2019	11,700	10,650	10,250	11,800	11,600	13,050	12,600	30,100	30,100	...	12,900
3	09/05/2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,100	30,100	...	12,850
4	09/06/2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,050	30,050	...	12,850
...											...	
98	January 17, 2020	10,800	10,450	11,900	11,700	13,100	12,650	33,100	33,100	118,500	...	122,050

3.2 Initial Data Exploration and Preprocessing

The initial data exploration and preprocessing stage was conducted to understand the structure and quality of food commodity price data prior to further analysis. This process involved examining the data volume, attributes, data types, missing values, duplicate data, as well as the consistency of price formats and observation periods. Next, non-compliant data were corrected or addressed, and the price data were converted to a numeric format and prepared for the stages of descriptive analysis, feature engineering, standardization, PCA, and K-Means clustering, as shown in Table 2.

Table 2. Data Exploration

Date	Rice	Low-Quality Rice I (kg)	Low-Grade Rice II (kg)	Medium-Quality Rice I (kg)	Medium-Grade Rice II (kg)	Super Grade I Rice (kg)	Super Grade II Rice (kg)	Chicken Meat	Fresh Breed Chicken Meat (kg)	Beef	...	Local Granulated Sugar (kg)
Sept 2, 2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,450	30,450	117,900	...	12,900
Sept 3, 2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,300	30,300	118,350	...	12,900
Sept 4, 2019	11,700	10,650	10,250	11,800	11,600	13,050	12,600	30,100	30,100	118,350	...	12,900
Sept 5, 2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,100	30,100	118,350	...	12,850
Sept 6, 2019	11,700	10,650	10,250	11,800	11,550	13,050	12,600	30,050	30,050	118,350	...	12,850

3.3 Feature Selection and Data Scaling

The feature selection and data scaling stages were carried out by selecting variables relevant to representing the price characteristics of each food commodity, such as average price, standard deviation, minimum and maximum values, price changes, and volatility. The selected features were then scaled using StandardScaler (Z-Score) so that each variable would have a comparable scale. This step is important because K-Means uses distance calculations to determine cluster membership; therefore, differences in scale between features can affect the clustering results in Table 3.

Table 3. Data after scaling:

Date	Rice	Low-Quality Rice I (kg)	Low-Grade Rice II (kg)	Medium-Quality Rice I (kg)	Medium-Grade Rice II (kg)	Super Grade I Rice (kg)	Super Grade II Rice (kg)	Chicken Meat	Fresh Breed Chicken Meat (kg)	Beef	..	Local Granulated Sugar (kg)
Sept	-	-	-	-	-	-	-	-	-	-	..	-0.697446

Date	Rice	Low-Quality Rice I (kg)	Low-Grade Rice II (kg)	Medium-Quality Rice I (kg)	Medium Grade II Rice (kg)	Super Grade I Rice (kg)	Super Grade II Rice (kg)	Chicken Meat	Fresh Breed Chicken Meat (kg)	Beef	Local Granulated Sugar (kg)
2, 2019	1.34183	0.97256	1.6002	0.36230	1.38547	1.04538	0.31801	2.14849	2.14849	1.22020	.
Sept 3, 2019	-	-	-	-	-	-	-	-	-	0.63560	..
Sept 4, 2019	1.34183	0.97256	1.6002	0.36230	1.38547	1.04538	0.31801	2.25707	2.25707	0.63560	..
Sept 5, 2019	-	-	-	-	-	-	-	-	-	0.63560	..
Sept 6, 2019	1.34183	0.97256	1.6002	0.36230	1.38547	1.04538	0.31801	2.40184	2.40184	0.63560	..
Sept 7, 2019	-	-	-	-	-	-	-	-	-	0.63560	..
Sept 8, 2019	1.34183	0.97256	1.6002	0.36230	1.38547	1.04538	0.31801	2.43804	2.43804	0.63560	..
Sept 9, 2019	-	-	-	-	-	-	-	-	-	0.63560	..

3.4 Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) stage is used to reduce the number of feature dimensions without losing important information in the data. PCA transforms features that have undergone standardization into a number of principal components capable of explaining most of the data variation. In this study, PCA was also used to facilitate the visualization of relationship patterns among commodities and to examine the data structure prior to the clustering process. The PCA results were subsequently used to support the analysis and interpretation of the K-Means Clustering results, as shown in Figure 2.

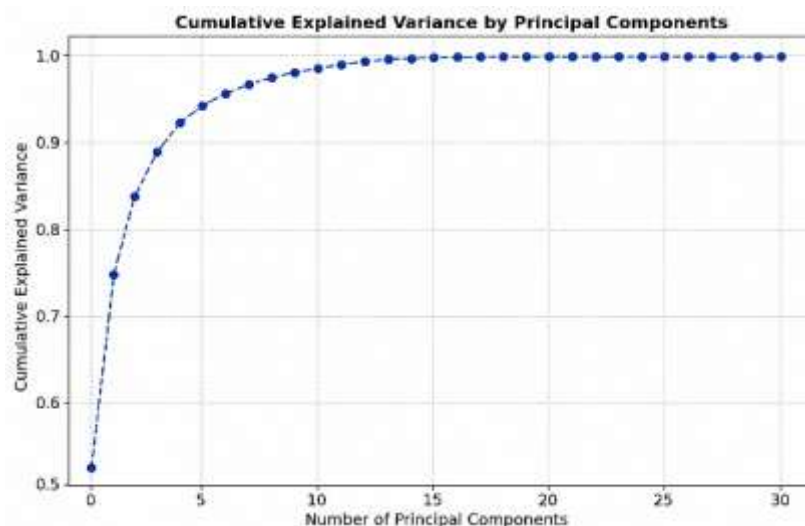


Figure 2. Principal Component Analysis (PCA)

Variance explained by each Principal Component:

PC1: 0.5200

PC2: 0.2290

PC3: 0.0867

PC4: 0.0536

PC5: 0.0291

PC6: 0.0164

PC7: 0.0136

PC8: 0.0112

PC9: 0.0075

PC10: 0.0068
 PC11: 0.0057
 PC12: 0.0045
 PC13: 0.0036
 PC14: 0.0034
 PC15: 0.0029
 PC16: 0.0018
 PC17: 0.0016
 PC18: 0.0011
 PC19: 0.0005
 PC20: 0.0003
 PC21: 0.0003
 PC22: 0.0002
 PC23: 0.0001
 PC24: 0.0001
 PC25: 0.0001
 PC26: 0.0000
 PC27: 0.0000
 PC28: 0.0000
 PC29: 0.0000
 PC30: 0.0000
 PC31: 0.0000

The PCA results show that PC1 is the most dominant component, contributing 52.00%, followed by PC2 at 22.90%. These two components account for 74.90% of the data's variation. When using five principal components, the total explained variance reaches 91.84%, meaning that most of the information in the dataset can be retained with far fewer dimensions than the original 31 features. Meanwhile, components beyond PC5 contribute relatively little to the data's variation. This indicates that PCA successfully performs dimensionality reduction effectively without losing most of the important information.

3.5 PCA Components and Data Transformation

Based on the cumulative explained variance results, the number of principal components was selected by considering each component's ability to retain information in the dataset. The analysis results show that the 10 principal components (PC1–PC10) account for 97.39% of the total data variance, meaning that only about 2.61% of the information is not represented by these ten components. Therefore, the 10 principal components were selected as a lower-dimensional representation of the original 31 features. Next, the data were transformed from the original feature space into a new space consisting of 10 principal components. This transformation produced a more concise representation of the data while retaining most of the information or variation present in the dataset. The data resulting from this PCA transformation can then be used for clustering analysis and visualization of food commodity group patterns in Table 4.

Table 4. Data after dimensionality reduction using 10 PCA components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Sept 2, 2019	-4.159944	3.367658	-	-	-	0.430493	1.749679	-0.898562	-0.569536	0.502383
Sept 3, 2019	-3.767640	4.033935	-	1.541027	-	-	0.889705	-0.468704	-0.274255	0.262420
Sept 4, 2019	-3.789873	4.010512	-	1.623473	-	0.459960	0.707708	-0.163937	-0.028471	-
Sept 5, 2019	-4.143481	3.779760	-	1.683336	-	0.062688	0.836764	-0.163129	-0.225991	0.123126
Sept 6, 2019	-4.191419	3.696245	-	1.671004	-	0.016059	0.747228	-0.071228	-0.134632	0.136429

3.6 PCA Results

The PCA results indicate substantial dimensional redundancy among the standardized price characteristics. PC1 explains 52.00% of the total variance, while PC2 explains 22.90%, resulting in a cumulative variance of 74.90%. The first five components explain approximately 91.84% of the total variance, while the first ten components increase the cumulative explained variance to 97.39%. Therefore, ten principal components were retained for subsequent clustering. This configuration substantially reduces the dimensionality while preserving most of the information contained in the original standardized feature space.

Because only a subset of principal components is retained, Euclidean distances in the reduced space represent an approximation of the distances in the original standardized space. The 97.39% explained variance indicates that the approximation retains most of the original information. Nevertheless, cluster interpretation is conducted using the original price characteristics to avoid interpreting PCA scores as direct economic indicators.

3.7 Determining the Optimal Number of Clusters Using the Elbow Method

Before applying the K-Means Clustering algorithm, the optimal number of clusters (k) is determined to obtain the most representative clustering structure. This determination is performed using the Elbow Method by evaluating the inertia value, or Within-Cluster Sum of Squares (WCSS), for several alternative numbers of clusters. The inertia value is calculated based on the total sum of the squared distances of each data point from its cluster centroid. Next, these values were visualized in a graph, and the “elbow” point was identified based on the point where the rate of decrease in the inertia value began to level off. The number of clusters at that point was considered the optimal value of k because it provides a balance between the level of compactness within clusters and the complexity of the number of clusters, as shown in Figure 3.

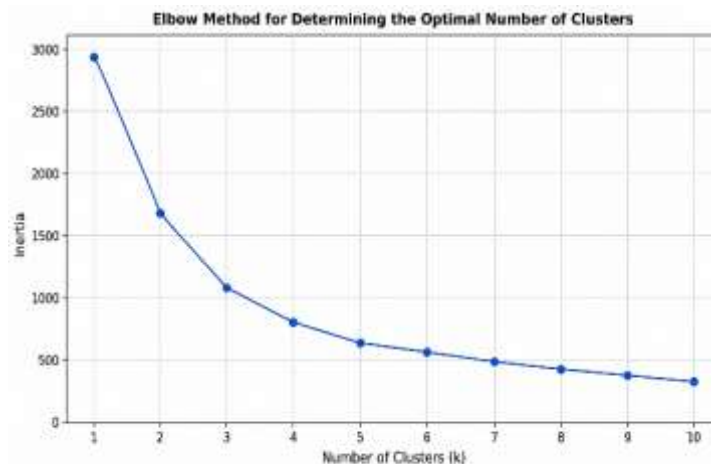


Figure 3. Optimal Number of Clusters Using the Elbow Method

3.8 Application of K-Means Clustering

After determining the optimal number of clusters—for example, $k=3$ based on the results of the Elbow Method—the K-Means Clustering algorithm is applied to group food commodities based on the similarity of their price characteristics. The clustering process is performed using data that has undergone standardization and PCA transformation. The K-Means algorithm iteratively determines the centroids, calculates the distance of each object from the centroids, assigns objects to the cluster with the closest distance, and updates the centroid positions until a stable clustering result is obtained. The final result of this process is a group of commodities with relatively similar price characteristics, which are then analyzed to identify patterns and differences in characteristics between clusters, as shown in Figure 4.

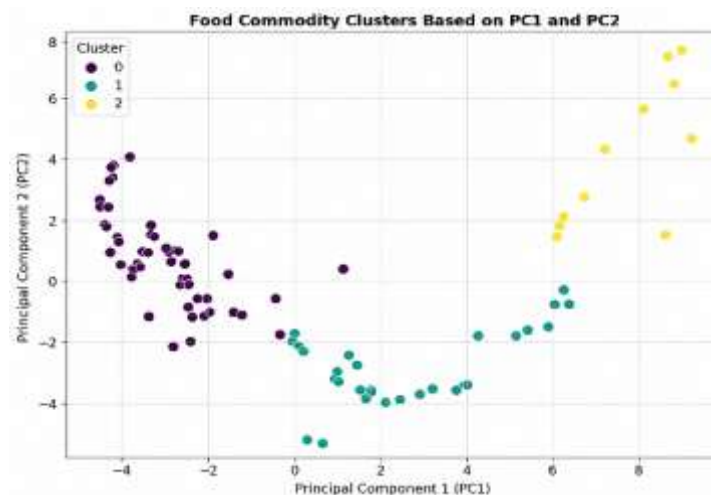


Figure 4. Visualization of Clustering Results

Once the optimal number of clusters has been determined based on the results of the Elbow Method—the K-Means Clustering algorithm is applied to group food commodities based on the similarity of their price characteristics. The clustering process is performed using data that has undergone standardization and PCA transformation. The K-Means algorithm iteratively determines the centroids, calculates the distance of each object from the centroids, assigns objects to the cluster with the closest distance, and updates the centroid positions until a stable clustering result is obtained. The final result of this process is a group of commodities with relatively similar price characteristics, which are then analyzed to identify patterns and differences in characteristics between clusters. K-Means was subsequently applied to the ten-dimensional PCA representation. Each commodity was assigned to the nearest centroid based on Euclidean distance. The resulting clusters were subsequently profiled using the original standardized price characteristics to preserve the interpretability of the clustering results.

3.9 Cluster Characteristic Analysis

The cluster characteristic analysis stage is conducted to identify and compare the characteristics of each commodity group generated by the K-Means algorithm. The analysis is based on the average values of each feature within each cluster, such as price level, standard deviation, coefficient of variation, price change, and volatility. The analysis results were used to determine the dominant characteristics of each cluster and to identify commodity groups with relatively stable price patterns, high price levels, or greater fluctuations. Subsequently, these characteristics served as the basis for interpreting the clustering results and explaining their implications for the price movement patterns of food commodities, as shown in Figure 5.



Figure 5. Cluster Characteristics

Based on the visualization of the standardized average features for each cluster, there are clear differences in price characteristics between Cluster 0, Cluster 1, and Cluster 2. The values on the heatmap represent the average of the standardized results; thus, positive values indicate characteristics above the overall average, while negative values indicate characteristics below the average.

Cluster 0 generally has negative values for most commodities. This indicates that commodities in this cluster tend to have price characteristics that are relatively lower than the overall data average. However, several chili pepper commodities show positive values, such as Red Bird's Eye Chili at 0.78, Green Bird's Eye Chili at 0.78, and Bird's Eye Chili at 0.74, indicating price characteristics that are relatively higher than the average.

Cluster 1 exhibits more diverse characteristics. Several commodities have positive values, particularly Breed Chicken Meat at 0.87, Chicken Eggs at 0.88, Fresh Breed Chicken Eggs at 0.88, and Red Onions at 0.83. Conversely, the chili pepper commodity group has fairly large negative values, ranging from approximately -1.22 to -1.31. This indicates that Cluster 1 has relatively moderate and heterogeneous price characteristics, with fairly strong differences between commodities.

Meanwhile, Cluster 2 shows predominantly positive values for most commodities. High values are observed for Branded Packaged Cooking Oil 2 at 2.37, Granulated Sugar at 2.32, Local Granulated Sugar at 2.27, Branded Packaged Cooking Oil 1 at 2.23, and Medium-Sized Garlic at 2.21. These conditions indicate that Cluster 2 is dominated by commodities with relatively high prices compared to the overall average.

Overall, the clustering results show that the three clusters have distinct characteristic patterns. Cluster 0 tends to represent commodities with relatively low price characteristics; Cluster 1 exhibits more moderate and heterogeneous price characteristics; while Cluster 2 is dominated by commodities with relatively high price characteristics. These differences indicate that the K-Means Clustering method is capable of identifying groups of

commodities based on the similarity of their price characteristics, and can therefore be used as a basis for understanding the grouping patterns and price differences among food commodities.

3.10 Discussion

The results demonstrate that food commodities in Indonesia exhibit heterogeneous multidimensional price characteristics, indicating that price levels alone are insufficient to describe commodity price behavior. PCA reduced the data dimensionality, with the first ten components retaining 97.39% of the total variance, consistent with its role in preserving information while simplifying multidimensional data [8], [9]

The clustering results produced three distinct groups: Cluster 0 represents relatively lower price characteristics, Cluster 1 exhibits moderate and heterogeneous characteristics, and Cluster 2 is dominated by relatively higher price characteristics. These differences indicate that commodity price behavior involves not only price levels but also variation, price changes, and volatility. The findings are consistent with previous studies on food commodity clustering and price volatility [5],[6] [17],[18].

Compared with previous studies focusing on volatility, individual commodities, or temporal price movements, this study contributes a broader classification by analyzing 31 food commodities using multiple price characteristics. The results complement previous research on inter-regional price spillovers and price transmission [19], [20]. Therefore, food price monitoring and stabilization should consider commodity-specific characteristics rather than applying uniform interventions. Overall, the PCA–K-Means approach provides a practical framework for identifying homogeneous commodity profiles and supporting differentiated price monitoring and stabilization strategies..

4. CONCLUSION

This study demonstrates that food commodities in Indonesia exhibit distinct multidimensional price profiles and therefore should not be evaluated solely based on average price levels. By integrating feature engineering, standardization, Principal Component Analysis (PCA), and K-Means clustering, the study reduced data dimensionality while retaining 97.39% of the total variance through ten principal components. The resulting three clusters represent distinct combinations of price level, dispersion, price changes, and volatility, demonstrating that commodities with similar price levels may have substantially different price-risk characteristics. The main contribution of this study is the development of a data-driven framework for profiling food commodities based on multiple price characteristics while maintaining interpretability through the original features. These findings provide practical support for differentiated food price monitoring and stabilization policies, particularly by enabling policymakers to prioritize commodities according to their respective price-risk profiles rather than applying uniform interventions. However, this study is limited by the relatively short observation period and its reliance on historical price-based indicators without incorporating supply, weather, distribution, inflation, and spatial-market factors. Consequently, the identified clusters should not be regarded as permanent classifications. Future research should employ longer observation periods, incorporate economic and spatial variables, and compare alternative clustering algorithms to assess the temporal stability and robustness of the identified commodity profiles..

5. DECLARATIONS

5.1 Author Contributions

Conceptualization: A.K. and A.P.J.;

Methodology: A.K.;

Software: A.K.;

Validation: Angga Putra Juledi

Formal Analysis: A.K. and A.P.J.;

Investigation: A.K. and A.P.J.;

Resources: A.K.;

Data Curation: A.P.J.;

Writing Original Draft Preparation: A.K. and A.P.J.;

Writing Review and Editing: A.K.;

Visualization: A.K.;

All authors have read and agreed to the published version of the manuscript.

5.2 Data Availability

The data presented in this study are available on request from the corresponding author.

5.3 Funding

Publication Fees for Collaborations Among Authors



5.4 Institutional Review Board Statement

Not applicable.

5.5 Informed Consent Statement

Not applicable.

5.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

5.7 Generative AI and AI-assisted Technologies in The Writing Process

The authors used Generative AI to improve the writing clarity of this paper. They reviewed and edited the AI-assisted content and took full responsibility for the final publication. Thankyou to those who have supported the implementation of this research.

REFERENCES

- [1] M. Azim, "Strategic Food Price Dynamics And Their Impact On Consumer Inflation In West Nusa Tenggara Province In 2024," *Journal of Tabulation Economic Data*, vol. 1, no. 1, Jun. 2025, doi: 10.67462/0QG5A524.
- [2] M. D. S. A. S. Alam, B. Rhama, F. Setiawan, A. I. U. Irwan, and F. Selvia, "The Indonesian Food Security Policy: Mapping of Traditional Markets in Central Kalimantan," *Policy & Governance Review*, vol. 8, no. 2, pp. 212–226, Jun. 2024, doi: 10.30589/PGR.V8I2.906.
- [3] J. Sosial Ekonomi dan Kebijakan Pertanian, M. Ridwansyah Matondang, and B. Krisnamurthi, "Price Fluctuations And Volatility Of National Strategic Food Commodities In Indonesia," *Agrisocionomics: Jurnal Sosial Ekonomi Pertanian*, vol. 8, no. 1, pp. 134–146, Mar. 2024, doi: 10.14710/AGRISOCIONOMICS.V8I1.17753.
- [4] S. Herliana, S. Ratnaningtyas, Q. Aina, U. Zuraida, A. Sutardi, and S. Qorina, "Supply and Demand of Rice in Indonesia: A Critical Review," *Journal of The Community Development in Asia*, vol. 8, no. 1, pp. 80–96, Jan. 2025, doi: 10.32535/JCDA.V8I1.3665.
- [5] 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.
- [6] M. A. Kesuma, C. A. Nugraha, O. I. Cahyani, M. Wati, and H. Haviluddin, "Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means," *JUKI: Jurnal Komputer dan Informatika*, vol. 7, no. 1, pp. 80–88, May 2025, doi: 10.53842/JUKI.V7I1.1012.
- [7] A. A. Nixia Tenriawaru, A. B. Razzaq Al Tawaha, R. C. Rukka, M. D. Ridwan, N. E. Varun Chan, and S. F. Hardiyanti Syam, "Vector Autoregressive (VAR) Method in Analyzing the Effect of Inflation on Food Price Volatility (FPV) in Palopo City, Indonesia," *International Journal of Professional Business Review*, vol. 8, no. 5, p. e01872, May 2023, doi: 10.26668/businessreview/2023.v8i5.1872.
- [8] W. R. Nurputra, W. Witanti, and A. Komarudin, "Principal Component Analysis (PCA) Untuk Meningkatkan Hasil Klasterisasi Penjualan Video Game Menggunakan Algoritma K-Means," *Jurnal Locus Penelitian dan Pengabdian*, vol. 4, no. 8, pp. 4683–4703, Aug. 2025, doi: 10.58344/LOCUS.V4I8.4151.
- [9] F. D. Winati, M. Arifin, M. I. Faturohman, and E. Mulyani, "Data-Driven Product Segmentation for Shallot Commodities using PCA and K-Means Clustering Approach," *JURNAL INFOTEL*, vol. 17, no. 3, pp. 490–503–490–503, Aug. 2025, doi: 10.20895/INFOTEL.V17I3.1307.
- [10] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," *Entropy 2021, Vol. 23, Page 759*, vol. 23, no. 6, p. 759, Jun. 2021, doi: 10.3390/E23060759.
- [11] M. Zmami, O. Ben-Salha, and O. Paper, "What factors contribute to the volatility of food prices? New global evidence," <https://agricecon.agriculturejournals.cz/doi/10.17221/99/2023-AGRICECON.html>, vol. 69, no. 5, pp. 171–184, 2023, doi: 10.17221/99/2023-AGRICECON.
- [12] N. Nairobi and L. Caroline, "Analysis of Food Inflation in Java and Sumatra Islands," *Indonesian Journal of Economics, Social, and Humanities*, vol. 3, no. 2, pp. 117–129, Aug. 2021, doi: 10.31258/IJESH.3.2.117-129.
- [13] M. Li, E. Frank, and B. Pfahringer, "Large scale K-means clustering using GPUs," *Data Mining and Knowledge Discovery* 2022 37:1, vol. 37, no. 1, pp. 67–109, Oct. 2022, doi: 10.1007/S10618-022-00869-6.
- [14] J. Sosial Ekonomi dan Kebijakan Pertanian, M. Ridwansyah Matondang, and B. Krisnamurthi, "Price Fluctuations And Volatility Of National Strategic Food Commodities In Indonesia," *Agrisocionomics: Jurnal Sosial Ekonomi Pertanian*, vol. 8, no. 1, pp. 134–146, Mar. 2024, doi: 10.14710/AGRISOCIONOMICS.V8I1.17753.
- [15] A. Punhani, N. Faujdar, K. K. Mishra, and M. Subramanian, "Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means," *IEEE Access*, vol. 10, pp. 115025–115032, 2022, doi: 10.1109/ACCESS.2022.3215568.
- [16] M. R. L. and N. Kulkarni, "Does the financialization of agricultural commodities impact food security? An empirical investigation," *Borsa Istanbul Review*, vol. 24, no. 2, pp. 280–291, Mar. 2024, doi: 10.1016/J.BIR.2024.01.001.
- [17] L. K. Sari, N. A. Achsani, B. Sartono, and L. Anggraeni, "Investigating The Asymmetric Effect of Food Commodity Price on The Volatility in Indonesia," *Jurnal Manajemen dan Agribisnis*, vol. 20, no. 3, pp. 440–440, Nov. 2023, doi: 10.17358/JMA.20.3.440.
- [18] M. R. Matondang, B. Krisnamurthi, and H. Herawati, "Price Fluctuations And Volatility Of National Strategic Food Commodities In Indonesia," *Agrisocionomics: Jurnal Sosial Ekonomi Pertanian*, vol. 8, no. 1, pp. 134–146, Mar. 2024, doi: 10.14710/AGRISOCIONOMICS.V8I1.17753.

- [19] A. Theresia, M. Ikhsan, F. N. Kacaribu, and S. Sumarto, “Spillover Effect of Food Producer Price Volatility in Indonesia,” *Economies* 2025, Vol. 13, Page 256, vol. 13, no. 9, p. 256, Sep. 2025, doi: 10.3390/ECONOMIES13090256.
- [20] S. Asrin, T. A. Putri, and A. D. Utami, “Transmisi Harga Beras di Indonesia Pada Masa Pandemi Covid-19,” *Jurnal Agribisnis Indonesia*, vol. 10, no. 1, pp. 159–168, Jun. 2022, doi: 10.29244/JAI.2022.10.1.159-168.
- [21] F. Badri, S. Umie, and R. Sari, “Penerapan Metode Principal Component Analysis (PCA) Untuk Identifikasi Faktor-Faktor yang Mempengaruhi Sikap Mahasiswa Memilih Melanjutkan Studi ke Kota Malang,” *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, pp. 426–431, Dec. 2021, doi: 10.47065/BITS.V3I3.1139.
- [22] Y. Dai, L. Yang, and Y. Cao, “Long baseline underwater source localization based on deep K-Means++ clustering in complex underwater environments,” *Digit. Signal Process.*, vol. 164, p. 105281, Sep. 2025, doi: 10.1016/J.DSP.2025.105281.
- [23] G. B. Kaligis and S. Yulianto, “Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai,” *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 1, no. 3, pp. 179–193, 2022, doi: 10.24246/itexplore.v1i3.2022.pp179-193.
- [24] 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.
- [25] J. Lu, T. Luo, and K. Li, “A forward k-means algorithm for regression clustering,” *Inf. Sci. (N. Y.)*, vol. 711, p. 122105, Sep. 2025, doi: 10.1016/J.INS.2025.122105.