Clustering and Forecasting Beef Prices: WPGMA and Time Series Regression Integration

Agrupamiento y Pronóstico de los Precios de la Carne de Res mediante WPGMA y Regresión en Series Temporales

Fachrian Bimantoro Putra [1*], Sri Wahyuningsih [1], Andrea Tri Rian Dani [1].

Highlights

- We propose an integrated time series clustering framework that enhances forecasting accuracy by capturing both temporal dynamics and structural grouping patterns in regional beef prices.
- The integration of PCA, WPGMA clustering, and time series regression enables streamlined analysis of price patterns with consideration for major religious calendar events.
- The proposed approach was applied to large-scaletime series dataset, demonstrating effective clustering and accurate forecasting of beef prices across Indonesian provinces.

Innovaciencia ISSN: 2346-075X

E- ISSN: 2346-075X
Innovaciencia2026; 14(1); e5312
http//dx.doi.org/10.15649/2346075X.5312

ORIGINAL RESEARCH:

How to cite this article:
Putra FB., Wahyuningsih S., Dani ATR. Clustering and forecasting beef prices: WPGMA and time series regression integration. *Innovaciencia* 2026; 14 (1); e5312.
http://dx.doi.org/10.15649/2346075X.5312

Received: 9 September 2025
Accepted: 15 December 2025
Published: 06 February 2026

Keywords:

Beef Price, Calendar Variation, PCA, TSR, WPGMA

Palabras clave:

precio de la carne de res, variación estacional, PCA, TSR, WPGMA.

ABSTRACT

Introduction. Beef is the second most consumed animal protein in Indonesia after chicken. Its demand increases during Eid al-Fitr and Eid al-Adha, driving seasonal price spikes. These fluctuations challenge the government in maintaining price stability, with regional differences further complicating policy decisions. **Objectives.** This study aims to group Indonesian provinces by monthly beef price trends using time series clustering and to perform group-level forecasting with Time Series Regression that incorporates the effects of Eid al-Fitr and Eid al-Adha. This approach simplifies modeling complexity and supports efficient policy-making for beef price stabilization. **Materials and Methods.** Principal Component Analysis (PCA) was used to optimize the dataset before clustering using the Weighted Pair Group Method with Arithmetic Mean (WPGMA) algorithm, combined with the Autocorrelation-Function (ACF) Distance similarity measure. Time Series Regression (TSR) models with dummy variables for Eid al-Fitr and Eid al-Adha effects were then applied to each cluster for forecasting. **Results.** The analysis identified five optimal clusters, with a silhouette coefficient value of 0.61. Forecasting within each cluster showed in-sample MAPE values ranging from 1.05% to 5.15%. All clusters exhibited an increasing trend in beef prices from September 2023 to August 2024, highlighting the impact of calendar events and regional characteristics on price dynamics. **Conclusions.** The clustering-based forecasting approach effectively simplifies price analysis across regions, supporting targeted policy interventions and improving the accuracy of beef price predictions in Indonesia.

RESUMEN

Introducción. La carne de res es la segunda proteína de origen animal más consumida en Indonesia, después del pollo. Su demanda aumenta durante las festividades de Eid al-Fitr y Eid al-Adha, lo que genera incrementos estacionales en los precios y representa un desafío para el mantenimiento de la estabilidad de precios, agravado por las diferencias regionales. **Objetivos.** Agrupar las provincias de Indonesia según las tendencias mensuales de los precios de la carne de res mediante clustering de series de tiempo y realizar pronósticos a nivel de grupo utilizando modelos de regresión de series de tiempo que incorporen los efectos de Eid al-Fitr y Eid al-Adha, con el fin de apoyar una formulación de políticas más eficiente. **Materiales y métodos.** Se aplicó Análisis de Componentes Principales para optimizar el conjunto de datos previo al agrupamiento mediante el algoritmo Weighted Pair Group Method with Arithmetic Mean (WPGMA), utilizando como medida de similitud la distancia basada en la Función de Autocorrelación. Posteriormente, se emplearon modelos de regresión de series de tiempo con variables dummy para capturar los efectos de las festividades en cada clúster. **Resultados.** Se identificaron cinco clústeres óptimos, con un coeficiente de silueta de 0,61. Los pronósticos presentaron valores de MAPE en la muestra entre 1,05 % y 5,15 %, y todos los clústeres mostraron una tendencia creciente de precios entre septiembre de 2023 y agosto de 2024. **Conclusiones.** El enfoque de pronóstico basado en agrupamiento simplifica el análisis regional de precios y contribuye a intervenciones de política pública más focalizadas y precisas para la estabilización del precio de la carne de res en Indonesia.

[1] Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University,
* Corresponding Author: putrafachrian9@gmail.com

http//dx.doi.org/10.15649/2346075X.5312

INTRODUCTION

Beef demand in Indonesia has risen due to changes in dietary preferences and growing public awareness of the nutritional value of animal protein, positioning beef as the second most consumed animal protein after poultry (1-2). The availability of meat is inversely proportional to the high demand, significantly ahead of major religious holidays such as Eid al-Fitr and Eid al-Adha. This results in high beef prices in the market. Various methods have been carried out to suppress the increase in beef prices. One of them is importing beef from Australia, the United States, and several other countries. However, this has not significantly impacted local beef prices, which continue to increase. Policies must consider the distinc characteristics of each región (3). Grouping provinces based on the attributes of beef prices and forecasting using models built from each group can make policy management more efficient because there is no need for different policies for each province.

Cluster analysis is the proper method to identify homogeneous objects into groups called clusters (4-5). The resulting clusters have high homogeneity for each object in one cluster and high heterogeneity between objects from other clusters (6-7). Cluster analysis can not only be used on cross-section data but can also be applied to time series data (8-9). A time series consists of sequential observations indexed over time, allowing for dynamic modeling and forecasting of trends or cycles (10-13). Beef price data in all provinces on the official website can be grouped.

There is a multicollinearity problem if grouping is done using time series data, which can result in poor cluster structures (14-15). Optimization methods are needed before grouping to overcome the problem of multicollinearity, one of which is by using PCA (16-17). PCA will reduce the number of variables into smaller dimensions so that one represents several studied variables (18).

Based on the background described, previous studies have applied clustering and forecasting methods, but most have not addressed the strong multicollinearity that arises in multi-regional time series data. This issue can distort similarity measurements among provinces, weaken cluster structures, and ultimately reduce the reliability of time series clustering outcomes. To overcome this limitation, this study introduces an optimization step using PCA prior to clustering. This approach reduces dimensionality, minimizes multicollinearity, and strengthens cluster validity, forming the key novelty of this research.

Motivated by this gap, the present study applies time series clustering to group Indonesian provinces based on monthly beef price movements. The objectives are to obtain optimal clusters of provinces using multivariate monthly beef price data, and to perform group-level forecasting using TSR that incorporates calendar variation effects from Eid al-Fitr and Eid al-Adha. Conducting forecasts at the cluster level reduces modeling complexity compared to building separate TSR models for each province, thereby improving the efficiency of policy formulation. The resulting clusters and their prototype-based forecasts are expected to support the government in designing more targeted and efficient strategies for stabilizing beef prices across regions.

MATERIALS AND METHODS

Principal Component Analysis

The variable reduction process is carried out by selecting principal components (PCs) that will replace a set of initial variables, where the correlation between predictor variables is eliminated by transforming the original predictor variables into new variables that are not correlated at all [19]. The selected PCs have the most significant variance, thus representing the entire set of initial variables. In practice, PCA is widely used because it can be applied to all data conditions without reducing the number of initial variables while addressing the problem of multicollinearity [20].

In this study, PCA was performed on standardized data using Z-scores, with normalization applied variable-wise rather than temporally. This ensures that each price variable contributes equally to the PCA transformation by removing differences in scale before computing the covariance and correlation matrices. In general, variable reduction using the PCA method is carried out in six stages: data standardization, calculating the variance-covariance matrix, calculating the correlation matrix, calculating eigenvalues and eigenvectors, determining the number of PCs formed, and forming new variables resulting from the reduction [19, 21]. Data standardization is carried out when the variables studied have different measurement scales. This process can be done by calculating the mean and standard deviation, known as Z-scores in Equation (1).

$$Z_{ip}^{*} = \frac{Z_{ip} - \bar{Z}_p}{S_p} \tag{1}$$

With:

Z_{ip}: time series data of the -th object and the -th variable

$\bar{Z}_p$: the mean of the -th variable's time series data

S_p: the standard deviation of the -th variable

Z_{ip}^{*}: standardization of the -th object's time series data for the -th variable

In the calculation of the variance-covariance matrix (Σx)(Σz), the standardized data will have the same values as the correlation matrix ($\mathbf{R}_z$). The covariance calculation is performed through Equation (2).

$$S_{Z_p Z_{p'}} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_{ip} - \bar{Z}_p)(Z_{ip'} - \bar{Z}_{p'}) \tag{2}$$

The calculation of the correlation matrix is performed through Equation (3)

$$\rho_{Z_p Z_{p'}} = \frac{S_{Z_p Z_{p'}}}{S_{Z_p} S_{Z_{p'}}} \tag{3}$$

Where $S_{Z_p Z_{p'}}$ is the covariance between variables Z_p and $Z_{p'}$.

A scalar λ is called an eigenvalue or characteristic value of a matrix A if there exists a nonzero vector x , such that $A\mathrm{x} = \lambda\mathrm{x}$. If matrix A is an $\eta_\eta \times \eta_\eta$ matrix, then the non-zero vector x in R^n is called an eigenvector of A. The determination of λ can be done using the determinant in Equation (4).

$$\det|\boldsymbol{\eta I} - \boldsymbol{R_z}| = 0 \tag{4}$$

After obtaining the eigenvalues and eigenvectors, the number of PCs was primarily determined using the Kaiser criterion, in which PCs with eigenvalues greater than 1 are retained. Eigenvalues indicate the amount of variance explained by each PC, where larger values reflect components that contribute more substantially to capturing the underlying structure of the data. To ensure that the selected components sufficiently represented the total variability, the cumulative explained variance percentage was also examined. This measure shows the proportion of total variance accounted for by the first several PCs and helps verify that the retained components collectively preserve the essential information contained in the original variables. In addition, the scree plot was inspected to identify the inflection point, which visually marks the transition from components with meaningful contributions to those dominated by noise. The interpretability of the resulting PCs was further supported by the factor loading table, which presents the correlations between the original variables and each PC. These loadings help clarify which variables predominantly shape each component. Taken together, the use of the Kaiser criterion, cumulative explained variance, scree plot assessment, and factor loading evaluation provides a coherent and evidence-based justification for the selected number of principal components. The new reduced-dimension variables were subsequently constructed through linear combinations as presented in Equation (5).

$$PC_{i,q} = \sum_{p=1}^{P} \boldsymbol{v}_p z_p \tag{5}$$

Where $\mathbf{v}_p$ is the eigenvector and q is the PC component.

Cluster Analysis

Cluster analysis is a method used to identify homogeneous objects into groups called clusters [22]. The resulting clusters have high homogeneity for each object within a cluster and high heterogeneity between objects from different clusters [6]. Hierarchical cluster analysis methods are generally divided into agglomerative and divisive methods [23,24]. Agglomerative methods rely on combining the most similar objects to progressively form a hierarchy of clusters. Some commonly used algorithms in this family include complete linkage, single linkage, average linkage, and the WPGMA [25]. In this study, WPGMA was selected because, despite being less frequently used in economic or commodity price studies, it offers several methodological advantages. First, WPGMA assigns equal weight to each cluster regardless of its size, preventing larger clusters from dominating the merging process. Second, it provides more balanced and stable cluster structures compared to single or complete linkage, which are prone to chaining effects or over-compact clusters. Accordingly, WPGMA has been widely applied in hierarchical analyses to explore structured relationships among variables in different research contexts [26].

Similarity Measurement

Autocorrelation-based distance is used to measure the distance between two time series objects [27]. This approach is preferred because it preserves the underlying time-dependent structure, ensuring that the dynamics across lags remain represented even when the series differ in scale or level. Autocorrelationbased distance offers a more stable and efficient alternative, particularly for large datasets. Based on the autocorrelation vectors resulting from estimating both time series data, the autocorrelation function can be determined in Equation (6).

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - Z)^2}, k = 0,1,2, \ldots \quad (6)$$

From this function, a formula can be determined to calculate the distance between time series data as follows

$$d_{ACF}\left(Z_{1(t)}, Z_{2(t)}\right) = \sqrt{\left(\hat{\mathrm{P}}_{Z_{1(t)}} - \hat{\mathrm{P}}_{Z_{2(t)}}\right)^T \mathbf{I} \left(\hat{\mathrm{P}}_{Z_{1(t)}} - \hat{\mathrm{P}}_{Z_{2(t)}}\right)} \quad (7)$$

where:

$d_{ACF}(Z_{1(t)}, Z_{2(t)})$: Autocorrelation distance of vector and

$\hat{\mathrm{P}}_{Z_{1(t)}}$: Estimation of autocorrelation vector

$\hat{\mathrm{P}}_{Z_{2(t)}}$: Estimation of autocorrelation vector

I : Identity Matrix

Clustering Using WPGMA

WPGMA has been applied as part of agglomerative hierarchical clustering approaches in different applied research contexts [28]. The WPGMA algorithm forms a rooted tree (dendrogram) that shows the structure of the pairwise distance matrix. WPGMA clustering begins with merging the two closest clusters, for example, U and V forming, and then continues with merging clusters with other clusters, such as W and V [23, 25]. The formula for calculating the arithmetic mean of cluster members can be written as follows.

$$d_{(U \cup V),W} = \frac{d_{U,W} + d_{V,W}}{2} \quad (8)$$

Silhouette Coefficient

The silhouette coefficient is one of the evaluation methods that can provide information consistent with the number of clusters in a dataset [29]. The silhouette coefficient value is calculated using the following formula.

$$SC(i) = \frac{b_i + a}{\max\{a_i, b_i\}} \tag{9}$$

where:

a_i : average distance of object i to all data in the same cluster

b_i : average distance of object i to all data in different clusters

A silhouette value close to +1 indicates well-separated and cohesive clusters, while values near 0 imply overlapping boundaries. Conversely, values approaching -1 indicate misclassified observations. The silhouette coefficient is the global silhouette coefficient, which is the average of each SC(i) [29,30].

TSR

TSR is one of the forecasting methods used to determine the influence of predictor variables conditioned at time on the response variable (Zt) for $t = 1,2,\ldots,n$ [31,32].The TSR model equation is shown in Equation (10).

$$Z_t = \beta_0 + \beta_1 X + e_t \tag{10}$$

where:

β_0, β_1 : Time series regression model parameters

Z_t : Response variable, time series data at time-t

X: Predictor variables ,$t = 1,2,\ldots,n$

ε: Error in time series data at time-t

The generalization of the TSR model with more than one predictor variable can be written as follows.

$$Z_t = \beta_0 + \beta_1 X_{1(t)} + \beta_2 X_{2(t)} + \cdots + \beta_v X_{v(t)} + e_t \tag{11}$$

TSR forecasting modeling allows the use of dummy variables in the form of intervention effects, calendar variation effects, and seasonal effects [33]. This model is shown through Equation (12). For clarity, an example of a dummy variable used in the TSR model is the Eid al-Fitr indicator, which equals 1 during months in which Eid al-Fitr occurs and 0 otherwise. A similar definition was applied for the Eid al-Adha dummy, enabling the model to capture holiday-related price variations.

$$Z_t = \beta_0 + \beta_1 D_{1,t} + \beta_2 D_{2,t} + \cdots + \beta_v D_{v,t} + e_t \tag{12}$$

The TSR parameters were estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residuals to obtain the estimator $\hat{\beta}$ [34,35].OLS was chosen because it provides unbiased and efficient estimates under the classical time-series regression assumptions, and it is widely applied for modeling the effects of intervention and calendar variation in short-term forecasting contexts.

Parameter Significance Testing

The testing of influential parameters is carried out simultaneously and partially. For the simultaneous test, the F test is used with the following hypotheses:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_v = 0$

H_1 : At least one $\beta_l \neq 0, l = 0,1,2, \dots, v$

The statistical test is given by:

$$Fvalue = \frac{SSR/v}{SSE/(n - v - 1)} \tag{13}$$

where:

$SSR = \sum_{t=1}^{n}(\hat{Z}_t - \bar{Z})^2$; $SSE = \sum_{t=1}^{n}(Z_t - \hat{Z}_t)^2$.

To further clarify, SSR and SSE together form the basis of the coefficient of determination R^2 and is the number of predictor variables. The decision rule for the F-test in this study is based on a comparison between the computed F-value and the critical F-value from the F-distribution table at a predetermined significance level (α). If the computed F-value is greater than the critical F-value, the null hypothesis (H_0) is rejected, indicating that at least one independent variable significantly influences the dependent variable [36]. As for the partial test, the t-test is used with the following hypotheses:

$H_0 : \beta_l = 0$

$H_1 : \beta_l \neq 0, l = 0,1,2, \dots, v$

With the following statistical test:

$$t_{hitung} = \frac{\hat{\beta}_l}{SE(\hat{\beta}_l)} \tag{14}$$

The null hypothesis is rejected if the test statistic exceeds the critical value [37].

Forecasting Accuracy

The Mean Absolute Percentage Error (MAPE) is one of the most widely used forecasting accuracy measures because it provides an interpretation in percentage terms, allowing readers to easily understand how close the predicted values are to the actual observations [38,39]. MAPE is calculated by comparing the absolute difference between the actual value and the forecast relative to the actual value, and then averaging these percentages. This metric is particularly appropriate for data with varying scales or when a measure is needed that enables comparison across different periods and objects. The calculation of the MAPE value is shown through Equation (15).

$$\text{MAPE} = \frac{1}{\text{n}}\sum_{t=1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right| \times 100\% \tag{15}$$

The accuracy of the forecasting results can be seen from the obtained MAPE value. The MAPE value is considered very accurate if less than 10 % [40].

Data Source

The data used in this study consists of beef price information obtained from the official website of the Strategic Food Price Information Center (PIHPS), published by Bank Indonesia [41]. This data covers the period from January 2018 to August 2023 and represents beef prices in traditional markets across Indonesia. The selection of this dataset follows a purposive sampling technique, where targeted selection based on relevance criteria. Traditional market prices are considered to better reflect consumers' actual purchasing conditions, unlike modern retail prices, which may be influenced by additional factors such as branding and promotions. By focusing on PIHPS data, this study ensures a comprehensive and standardized source that facilitates accurate price trend analysis across regions.

The purposive sampling approach is particularly suitable for this study as it allows the selection of data that aligns with the need to analyze beef price fluctuations and their economic implications. The chosen timeframe January 2018 to August 2023 captures multiple economic cycles, including seasonal price spikes during Eid al-Fitr and Eid al-Adha and potential market disruptions caused by external factors such as policy changes or supply chain issues. By systematically selecting data from a reliable and official source, this study ensures the validity of the analysis while enabling effective clustering and forecasting of beef prices in different provinces.

RESULTS

Clustering with WPGMA

Before conducting the main analysis, an initial data exploration is carried out to understand the characteristics and trends of beef prices over time. One of the key steps in this exploration is visualizing the data through time series plots. For instance, (Figure 1) presents the time series plot of beef prices in East Kalimantan from January 2018 to August 2023, expressed in Indonesian Rupiah per kilogram (IDR/kg), where the monthly price movements exhibit an upward trend from January 2018 to August 2023. This increasing pattern suggests an upward tendency in beef prices over the years, which may be influenced by factors such as inflation, supplydemand dynamics, and seasonal variations, particularly around Eid al-Fitr and Eid al-Adha.

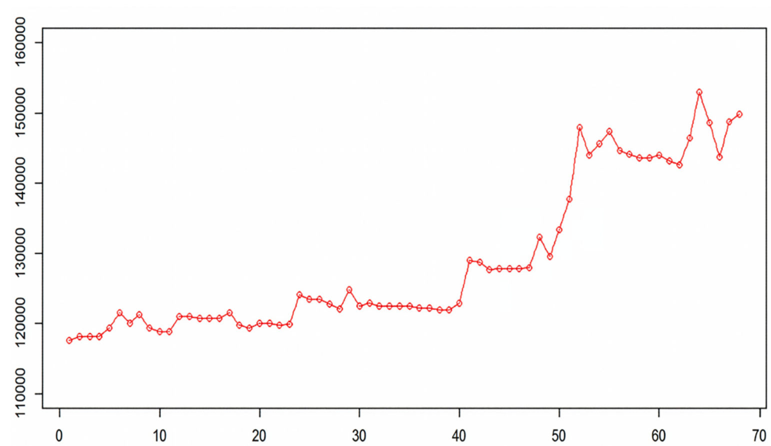

Figure 1. Time series plot of beef prices in East Kalimantan

To further understand price movements across different provinces, spatial mapping is utilized to visualize regional variations in beef prices. This approach allows for a more precise comparison of price differences and regional trends, which is crucial for identifying clusters of provinces with similar price patterns. The spatial mapping.gif provides an interactive visualization of provincial beef price variations across Indonesia, allowing readers to observe how price fluctuations are distributed geographically. To facilitate easy access, this visualization is made available through a QR code included below, which directs readers to the supplementary material where the animation can be viewed. This additional resource helps enhance the interpretation of regional price dynamics before proceeding to the clustering and forecasting analyses.

a

b

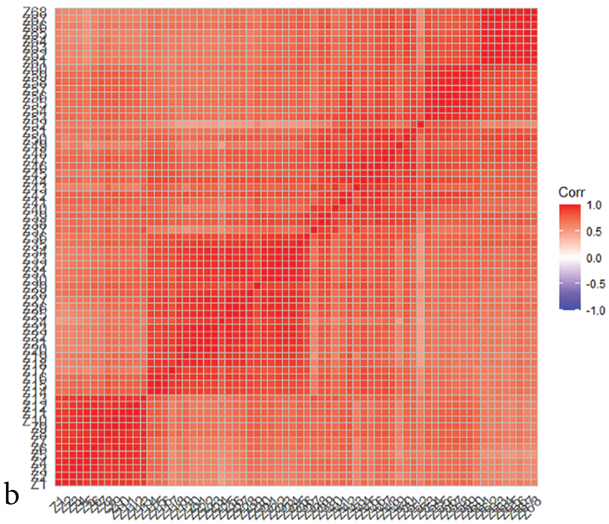

Figure 2. Spatial distribution and a heatmap matrix of beef prices in East Kalimantan. (a) Barcode spatial mapping, (b) Heatmap matrix

A dynamic visualization illustrating monthly beef price fluctuations across Indonesian provinces is presented in **(Figure 2a)**. Color transitions highlight how price levels vary regionally over time, providing an intuitive overview of spatial price dynamics. Additional details on the visualization process are included in the supplementary materials. Each province is categorized into one of three price levels: Low, Medium, and High, making it easier to identify areas with consistently high beef prices and regions where prices remain relatively stable. This visual representation highlights how beef prices are not uniform across Indonesia, influenced by regional supply chains, transportation costs, and local demand dynamics. A noticeable pattern in the visualization reveals that certain provinces experience a continuous upward trend in beef prices, indicated by gradual color shifts over time, as seen in one example province, East Kalimantan. This trend suggests that beef prices in Indonesia tend to rise month by month, reinforcing concerns about price stability and food affordability. Understanding these regional variations is crucial for policymakers in developing targeted strategies to manage price fluctuations, ensure market stability, and anticipate potential economic impacts, especially around festive seasons like Eid al-Fitr and Eid al-Adha when demand for beef surges significantly.

Multicollinearity detection will be performed on the monthly variables of the beef price data using the Pearson correlation matrix, which is visualized through a heatmap shown in **(Figure 2b)**. The intensity of the red color indicates strong positive correlations between variables. The presence of darker red blocks across the matrix indicates that beef prices over different months are highly correlated, suggesting a consistent pattern in price fluctuations over time. Although such high temporal correlations are typical in monthly time series data and may not be problematic for models that apply differencing or detrending, they can distort similarity measurements in clustering. Therefore, addressing multicollinearity through dimensionality reduction is necessary before performing time series clustering.

This strong correlation is a key indicator of multicollinearity, which implies that the price in one month is exhibiting strong serial correlation across consecutive time periods on previous months. Such a pattern is expected in time series data, where trends and seasonal effects often influence price movements. To mitigate this multicollinearity and extract orthogonal components that better represent the underlying structure of the data, PCA is applied prior to clustering. The multicollinearity issue in the time series data will be addressed using the PCA method with the following steps:

Data standardization

To bring the data to the same range, data standardization is performed using Z-score standardization, which will make the average value 0 and the standard deviation 1.

Variance-covariance matrix and correlation matrix

The variance-covariance matrix and the correlation matrix will have the same values because the data used has been standardized. These two matrices will serve as the basis for calculating eigenvalues.

Calculating eigenvalues and eigenvector

In this study, the PCA procedure was applied to a dataset consisting of 34 provinces and 68 monthly variables, resulting in a data matrix of size 34×68. Prior to PCA, all variables were standardized using Z-scores to ensure comparability across months. Because PCA was performed for the purpose of reducing the number of monthly variables, the analysis was conducted on the 68×68 correlation matrix derived from the standardized data. The eigen-decomposition of this matrix produces 68 eigenvalues and 68 corresponding eigenvectors, where each eigenvector has a dimension of 68×1. This size reflects the fact that each eigenvector represents a set of weights applied to the 68 monthly variables to form a principal component. The principal component scores for the 34 provinces are then obtained by multiplying the standardized data matrix with the selected eigenvectors, resulting in a reduced-dimension representation used for subsequent clustering analysis.

Determining the number of PCs formed

The number of PCs formed is determined by eigenvalues with a value ≥ 1 and is expected to explain more than 80% of the variance in the data. Therefore, in this study, five eigenvalues met these criteria, namel $\eta_1 = 5{,}30, \eta_2 = 4{,}59, \eta3 = 4{,}05, \eta4 = 1{,}65$, dan $\eta5 = 1{,}23$.

Forming new reduced variables

Based on the results of the eigenvalue and eigenvector calculations, the PC equations formed to construct the new reduced PCA variables will be obtained.

Since there are five eigenvalues greater than 1, this study identifies five PCs as the optimal number for dimensionality reduction. This selection ensures that the retained components capture the most significant variance in the data, as illustrated in **(Figure 3)**.

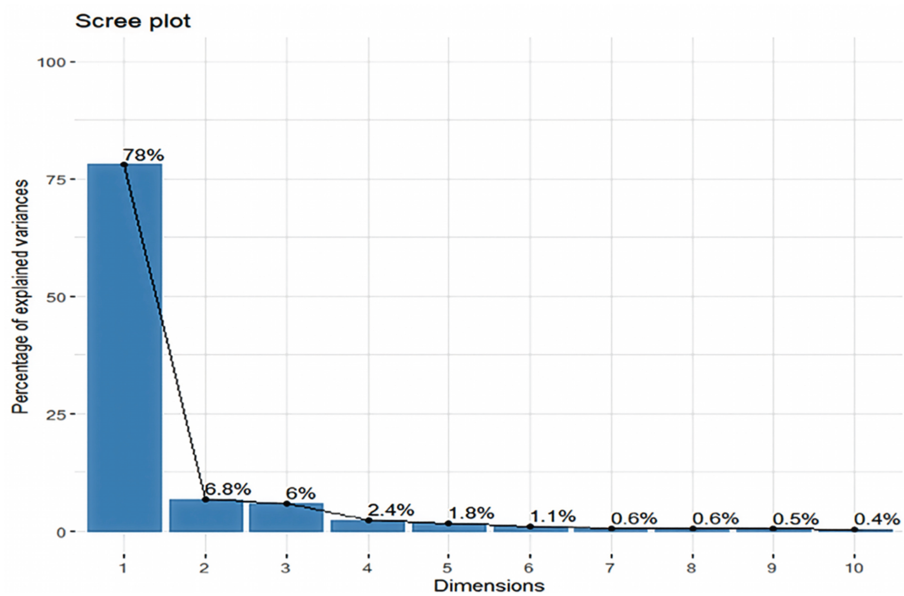

Figure 3. Scree Plot PCA

The clustering process utilizes the five previously formed PCs as input variables. To measure similarity between provinces, the ACF distance is applied, while the WPGMA serves as the clustering algorithm. The results of the ACF distance similarity distance calculation are shown in matrix D in Equation (16).

$$D = \begin{array}{c|cccccccc} & Z(1) & Z(2) & Z(3) & Z(4) & Z(5) & \cdots & Z(12) & \cdots & Z(34) \\ Z(1) & 0,00 & 0,26 & 0,23 & 0,57 & 0,78 & \cdots & 0,04 & \cdots & 0,82 \\ Z(2) & 0,26 & 0,00 & 0,46 & 0,80 & 0,30 & \cdots & 0,30 & \cdots & 1,04 \\ Z(3) & 0,23 & 0,46 & 0,00 & 0,56 & 0,20 & \cdots & 0,20 & \cdots & 0,73 \\ Z(4) & 0,57 & 0,80 & 0,56 & 0,00 & 0,56 & \cdots & 0,54 & \cdots & 0,55 \\ Z(5) & 0,78 & 0,30 & 0,20 & 0,56 & 0,00 & \cdots & 0,75 & \cdots & 0,11 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ Z(12) & 0,04 & 0,30 & 0,20 & 0,54 & 0,75 & \cdots & 0,00 & \cdots & 0,81 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ Z(34) & 0,82 & 1,04 & 0,73 & 0,55 & 0,11 & \cdots & 0,81 & \cdots & 0,00 \end{array} \quad (16)$$

The first pair of provinces merged were Aceh and West Java, based on minimum inter-cluster distance, after which the distance will be updated using Equation (8). This process continues iteratively, with clusters gradually forming and the distance matrix updating at each step until all provinces are grouped into a single large cluster. The clustering results in Figure 4 illustrate the hierarchical structure of regional similarities in beef price trends. As new clusters emerge, they reveal underlying patterns in price fluctuations, potentially influenced by factors such as geographic proximity, economic similarities, or shared market dynamics. The final clustering structure provides valuable insights for policymakers, enabling them to categorize provinces based on their pricing behavior and tailor strategies accordingly to ensure more effective price stabilization efforts.

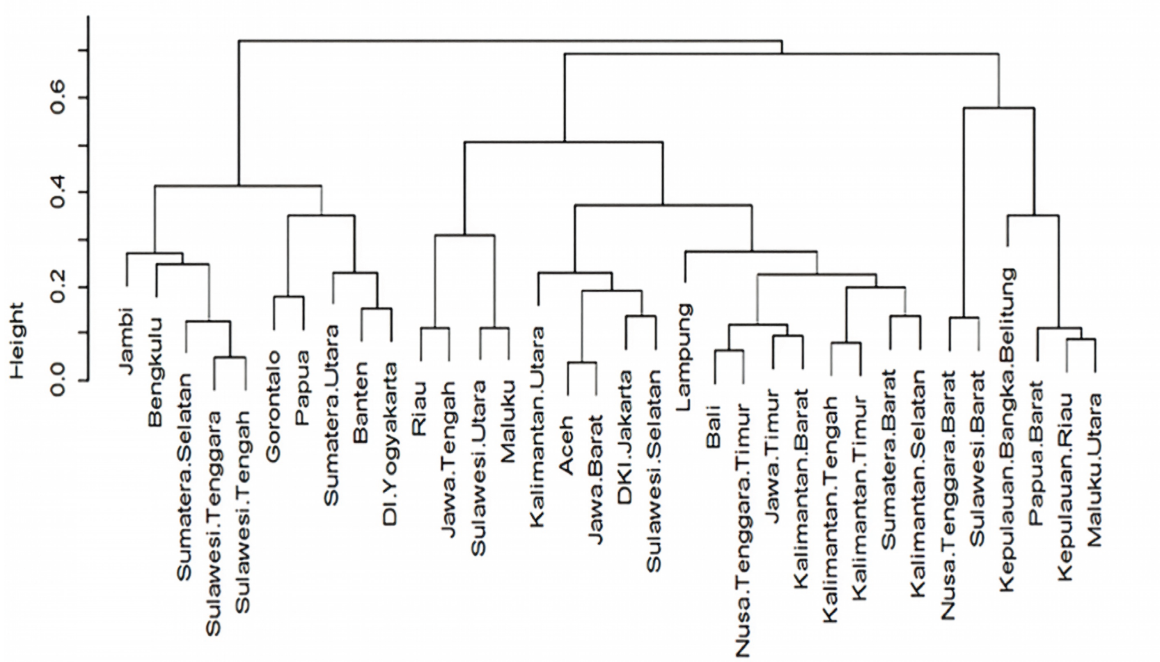

Figure 4. WPGMA clustering dendogram

The optimal number of clusters is determined using the silhouette coefficient. This metric evaluates the clustering quality by measuring how well each object fits within its assigned cluster compared to other clusters. This coefficient considers the cohesion within a cluster (the average distance between an object and other objects in the same cluster) and the separation from different clusters (the average distance between an object and the nearest neighboring cluster). A higher silhouette coefficient indicates a better-defined clustering structure, suggesting that the objects are well-grouped. A graph of the silhouette coefficient values for different cluster numbers, used to visualize the effectiveness of the clustering results, is presented in **(Figure 5)**.

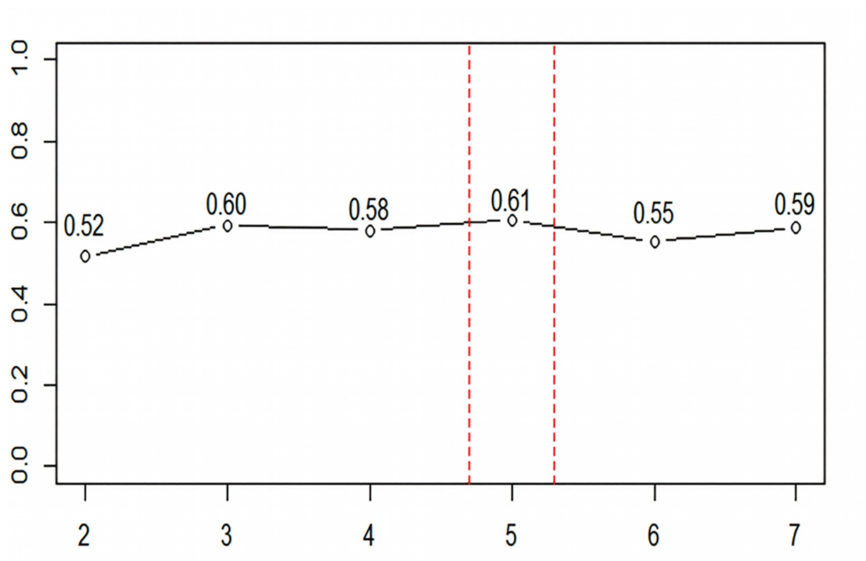

Figure 5. Silhouette coefficient value

Based on the silhouette coefficient analysis, the optimal number of clusters is determined to be five, with a silhouette coefficient value of 0.61. While other cluster configurations exhibit relatively close silhouette values, A five-cluster solution yielded the highest average silhouette value (0.61), indicating an optimal balance between cohesion and separation. This suggests that the five-cluster solution effectively balances intra-cluster cohesion and inter-cluster separation, making it the most suitable choice for categorizing beef price patterns across provinces. The final clustering results, detailing the distribution of provinces within each cluster, are presented in **(Table 1)**, providing a clear framework for further analysis and policy considerations.

Table 1. Optimal Cluster Grouping Results

Cluster	Province
1	Aceh, Sumatera Barat, Lampung, Jawa Barat, DKI Jakarta, Jawa Timur, Bali, NTT, Kali-man-tan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Selatan
2	Sumatera Utara, Jambi, Bengkulu, Sumatera Selatan, Banten, DI Yogyakarta, Gorontalo, Su-lawesi Tenggara, Sulawesi Tengah, Papua
3	Riau, Jawa Tengah, Sulawesi Utara, Maluku
4	Kep. Riau, Kep. Bangka Belitung, Maluku Utara, Papua Barat
5	Nusa Tenggara Barat, Sulawesi Barat

TSR Modelling

This study employs an 85% in-sample and 15% out-sample data split for model development. A total of five TSR models will be constructed, with one model assigned to each cluster.

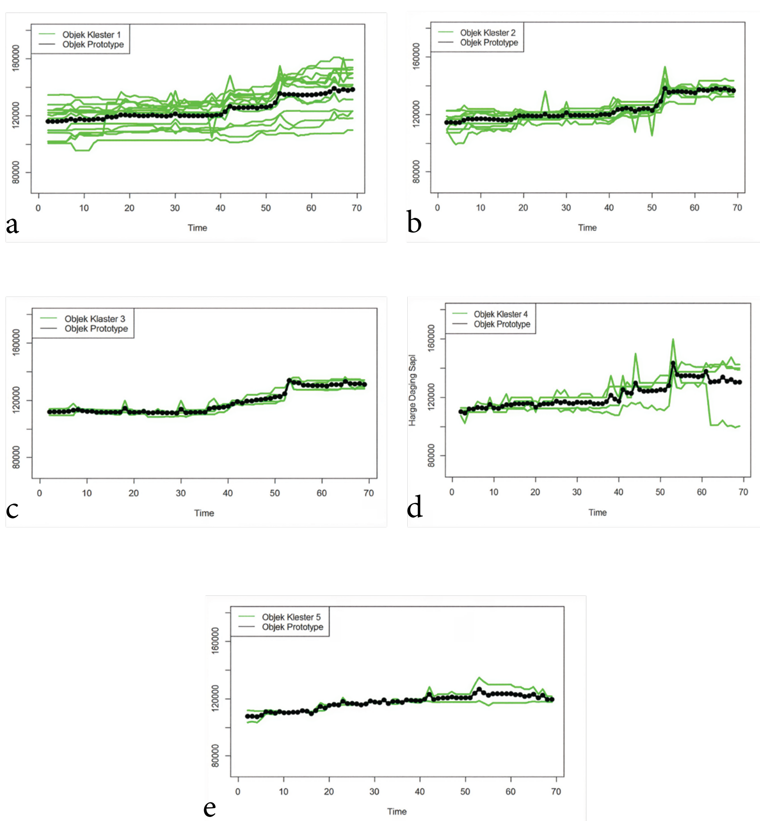

Figure 6. Multiple Time Series Plot. (a) cluster 1, (b) cluster 2, (c)cluster 3, (d) cluster 4, (e)cluster 5

Based on **(Figure 5)**, these prototype objects are determined based on the average beef price within each cluster, serving as the foundation for TSR model formation. The variables used in the TSR modeling are detailed as follows.

$$D_{1,t} = \begin{cases} 1, one\ month\ before\ Eid\ al-Fitr \\ 0, other \end{cases}$$

$$D_{2,t} = \begin{cases} 1, the\ month\ of\ Eid\ al-Fitr \\ 0, other \end{cases}$$

$$D_{3,t} = \begin{cases} 1, one\ month\ after\ Eid\ al-Fitr \\ 0, other \end{cases}$$

$$D_{4,t} = \begin{cases} 1, one\ month\ before\ Eid\ al-Adha \\ 0, other \end{cases}$$

$$D_{5,t} = \begin{cases} 1, the\ month\ of\ Eid\ al-Adha \\ 0, other \end{cases}$$

$$D_{6,t} = \begin{cases} 1, one\ month\ after\ Eid\ al-Adha \\ 0, other \end{cases}$$

Variables D3 and D4 occur in the same month, thus causing a singularity problem. In this study, one of them, namely D3, will be used, one month after Eid al-Fitr. The estimation of the TSR model parameters uses Ordinary Least Square (OLS), so the TSR model for each cluster is obtained as follows:

Cluster 1

$$\hat{Z}_t = 113.050{,}18 + 293{,}33T_t + 1.835{,}74D_{1,t} + 2.637{,}43D_{2.t} + 1.485{,}10D_{3,t} + 1.404{,}57D_{5,t} + 770{,}00D_{6.t}$$

Cluster 2

$$\hat{Z}_t = 111.311{,}53 + 313{,}97T_t + 2.358{,}62D_{1,t} + 2.839{,}64D_{2.t} + 1.994{,}67D_{3,t} + 1.834{,}99D_{5,t} + 1.203{,}52D_{6,t}$$

Cluster 3

$$\hat{Z}_t = 106.262{,}29 + 321{,}85T_t + 2.377{,}23D_{1,t} + 2.365{,}38D_{2.t} + 1.323{,}53D_{3,t} + 804{,}00D_{5,t} + 613{,}40D_{6,t}$$

Cluster 4

$$\hat{Z}_t = 107.155{,}04 + 397{,}43T_t + 4.265{,}57D_{1,t} + 1.893{,}15D_{2.t} + 605{,}72D_{3,t} + 2.568{,}98D_{5,t} + 609{,}06D_{6,t}$$

Cluster 5

$$\hat{Z}_t = 107.900 + 283{,}8T_t + 2.024D_{1,t} + 1.615D_{2.t} + 441{,}70D_{3,t} + 1.188D_{5,t} + 23{,}27D_{6,t}$$

The results of the simultaneous parameter significance test for each TSR model are shown in Table 2.

Table 2. Simultaneous Test of TSR Model

Cluster	Pvalue	Decision
1	$2.2 \times 10_{-16}$	Rejected H_0
2	$2.2 \times 10_{-16}$	Rejected H_0
3	$2.2 \times 10_{-16}$	Rejected H_0
4	$2.2 \times 10_{-16}$	Rejected H_0
5	$2.2 \times 10_{-16}$	Rejected H_0

The results of the simultaneous test for the TSR models across all clusters are presented in **(Table 2)**. The p-values for each cluster are consistently 2.2×10^{-16}, which is significantly lower than the 0.05 threshold.. This indicates that the null hypothesis (H_0), which states that all regression coefficients are equal to zero, is rejected in every cluster. In other words, the dummy variables collectively have a statistically significant impact on beef prices in Indonesia. To further examine the influence of each individual variable, a partial parameter significance test is conducted, with the results summarized in **(Table 3)**.

Table 3. Partial Test Results

Parameter	Variable	Cluster				
		1	2	3	4	5
	Intercept	0.00*	0.00*	0.00*	0.00*	0.00*
	T	0.00	0.00*	0.00*	0.00*	0.00*
	D1	0.14	0.12	0.22	0.01*	0.00*
	D2	0.03*	0.06	0.22	0.28	0.00*
	D3	0.23	0.19	0.49	0.72	0.46
	D5	0.26	0.23	0.67	0.14	0.05
	D6	0.53	0.43	0.75	0.72	0.96

Note: (*) significant

The results of the partial significance test for each variable in the TSR models across all clusters are presented in **(Table 3)**. The T variable, representing time, is consistently significant in nearly all clusters, indicating a strong time-dependent trend in beef price movements. This suggests that beef prices exhibit a clear pattern over time, reinforcing the importance of time as a key predictor in the model. Additionally, D1 and D2, which correspond to the month before Eid al-Fitr and the month of Eid al-Fitr, show a significant effect in clusters 1, 4, and 5. This implies that in these regions, the festive season has a measurable impact on beef prices, likely due to increased demand. However, in clusters 2 and 3, these variables are not statistically significant, suggesting that other factors, such as local market structures or alternative supply chain conditions, might influence the price dynamics in these regions. Meanwhile, D3, D5, and D6 are generally insignificant across most clusters, indicating that the months following Eid have a weaker or inconsistent effect on beef price fluctuations. These findings highlight the necessity of considering regional variations when implementing pricing and market intervention policies.

The forecasting accuracy evaluation using the MAPE value based on Equation (15), the MAPE values for each cluster for both in-sample and out-sample data are presented in **(Table 4)**.

Table 4. MAPE Accuracy Value

	MAPE	
Cluster	In-Sample	Out-Sample
1	1.55%	3.37%
2	2.04%	3.58%
3	5.15%	0.90%
4	3.37%	1.89%
5	4.60%	8.02%

The Mean Absolute Percentage Error (MAPE) values for both in-sample and out-sample data across all clusters are presented in **(Table 4)**. The relatively low MAPE values indicate that the TSR models demonstrate substantial predictive accuracy in forecasting beef prices. Most clusters exhibit in-sample MAPE values below 5%, with cluster 1 achieving the lowest error at 1.55%. This suggests that the models fit the historical data well. Cluster 3 shows the best predictive performance for out-sample accuracy with a MAPE of 0.90%, indicating minimal deviation from actual values. However, cluster 5 has the highest out-sample MAPE at 8.02%, suggesting greater volatility or unaccounted factors influencing beef prices in this group. Despite these variations, the TSR models remain reliable tools for predicting beef price trends in Indonesia from September 2023 to August 2024. The results of beef price forecasting in Indonesia for the next 12 periods are visualized using a time series graph in **(Figure 7)**.

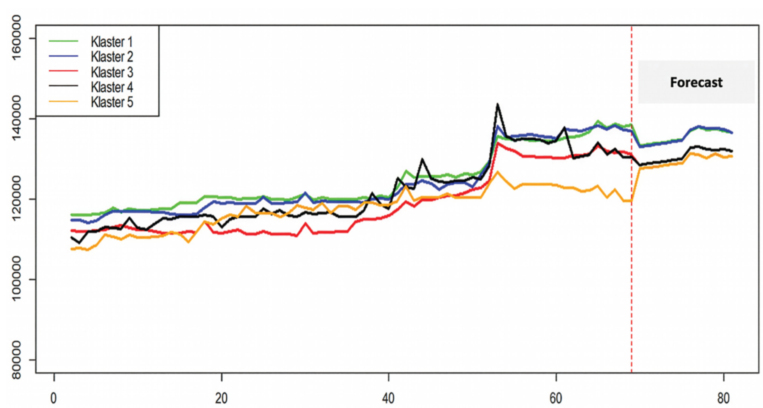

Figure 7. Beef price forecasting results

Based on **(Figure 8)**, beef prices are experiencing an increase, as indicated by the upward trend in the time series graph. Understanding this pattern is crucial for policymakers in formulating strategies to stabilize prices, manage supply chains, and anticipate market fluctuations. The TSR model used in this study serves as an effective approach to forecasting beef price movements. By incorporating key variables, including calendar variations, TSR provides a structured and data-driven method to predict future price trends. This modeling framework provides useful insights for policymakers and agricultural market regulators.

DISCUSSION

The results demonstrate that the integration of Principal Component Analysis (PCA), Weighted Pair Group Method with Arithmetic Mean (WPGMA), and Time Series Regression (TSR) provides an effective framework for clustering and forecasting regional beef prices in Indonesia. The optimal five-cluster solution, supported by a silhouette coefficient of 0.61, indicates a well-defined clustering structure with satisfactory cohesion and separation [29]. Beyond confirming methodological adequacy, these clusters reflect the existence of distinct regional price regimes, indicating that beef markets in Indonesia operate under heterogeneous structural and demand conditions rather than a single national pattern [42,43].

The application of PCA prior to clustering proved essential in addressing the strong multicollinearity observed among monthly beef price variables. High intercorrelations across time periods are common in economic time series, yet they can distort similarity measurements and weaken cluster validity. By reducing the original 68 monthly variables into five orthogonal principal components, this study preserved most of the data variability while minimizing redundancy. This supports earlier studies that emphasize the importance of dimensionality reduction when clustering highly correlated time series data [14,20].

The selection of WPGMA as the hierarchical clustering method also contributed to stable and balanced cluster structures. Unlike single or complete linkage methods that are sensitive to chaining effects or overly compact clusters, WPGMA assigns equal weight to each cluster during the merging process, preventing dominance by large clusters. The resulting clusters reveal that provinces with similar beef price dynamics are not always geographically adjacent, suggesting that price behavior is driven more by market integration, supply chain structure, and consumption patterns than by spatial proximity alone. This supports previous evidence that regional beef markets in Indonesia are interconnected and influenced by national-level supply and import policies [1,2].

The TSR models developed at the cluster level achieved high forecasting accuracy, with in-sample MAPE values ranging from 1.55% to 5.15% and out-sample values between 0.90% and 8.02%. These results confirm that cluster-based forecasting is more efficient and scalable than constructing separate models for each province while maintaining strong predictive performance. The consistent significance of the time variable across nearly all clusters indicates a persistent upward trend in beef prices. This trend likely reflects structural pressures such as increasing demand, limited domestic production capacity, and continued reliance on imports, highlighting the need for long-term supply-side policy responses [44]. Moreover, the significance of the Eid al-Fitr dummy variables (D1 and D2) in several clusters highlights the role of seasonal demand shocks during religious holidays. However, their insignificance in other clusters suggests heterogeneous regional responses, reinforcing the need for region-specific policy interventions rather than uniform national strategies.

In summary, this study confirms that integrating PCA, WPGMA clustering, and Time Series Regression provides an efficient framework for grouping and forecasting regional beef prices in Indonesia. The approach successfully addresses multicollinearity, produces well-defined clusters, and delivers accurate forecasts while reducing modeling complexity. Moreover, the findings highlight the importance of accounting for regional heterogeneity and calendar effects, offering a practical and data-driven basis for designing targeted price stabilization policies. From a policy perspective, this framework enables more targeted interventions by identifying clusters that are more vulnerable to price shocks or seasonal volatility, allowing resources to be allocated more efficiently [45].

Despite these contributions, several limitations must be acknowledged. First, the analysis relies solely on historical price data and does not explicitly incorporate supply, demand, or logistical variables such as production, imports, transportation costs, or consumption patterns. Second, the linear TSR framework may not fully reflect nonlinear or asymmetric price dynamics. Future studies could extend this framework by integrating hybrid machine learning models or incorporating regional economic indicators to further improve predictive accuracy and policy relevance.

CONCLUSIONS

This study demonstrates that the use of PCA as an optimization method successfully addresses multicollinearity issues and enhances the clustering structure. The clustering process, conducted using the WPGMA algorithm with ACF distance similarity, identifies an optimal cluster count of five, achieving a silhouette coefficient value of 0.61. The model achieved an average MAPE of 2.5% in-sample and 3.8% out-sample. These results indicate that regional characteristics significantly influence beef price dynamics, highlighting the importance of cluster-based analysis for policy formulation, including measures such as price stabilization and targeted subsidies. By grouping provinces with similar price patterns, policymakers can design more efficient and region-specific interventions to manage beef market fluctuations across Indonesia.

The beef price forecasting results reveal a consistent upward trend for the next 12 periods, from September 2023 to August 2024. This indicates the necessity for the government to proactively implement policies that regulate beef prices, ensuring stability in the market while safeguarding consumer purchasing power. Timely interventions can also help mitigate price volatility, particularly during periods of high demand such as religious holidays. A limitation of this study is the exclusion of exogenous macroeconomic variables, such as inflation, fuel prices, and beef import volumes, which may also affect regional price fluctuations. Future studies could extend this framework by integrating hybrid machine learning models or incorporating regional economic indicators to further improve predictive accuracy and policy relevance.

FUNDING

No external funding was received for this research. The study was carried out using the authors' own resources and the institutional facilities of the Faculty of Mathematics and Natural Sciences, Universitas Mulawarman, where the research was conducted.

ETHICAL CONSIDERATIONS

This study used publicly available secondary beef price data obtained from the Strategic Food Price Information Center (PIHPS) published by Bank Indonesia, covering the period from January 2018 to August 2023. No human participants, animals, or personal data were involved; therefore, ethical approval and informed consent were not required.

DECLARATION OF COMPETING INTEREST

There are no competing interests to declare.

REFERENCES

1. Firmansyah, Maruli P, Harahap A. Analysis of beef market integration between consumer and producer regions in Indonesia. *Open Agric* 2023;8. https://doi.org/10.1515/opag-2022-022

2. Kusumaningrum R, Darjanto A, Nurmalina R, Mulatsih S, Suprehatin. Effect of Import Policy on Beef Supply and Demand in Indonesia before and after the COVID-19 Pandemic. *Tropical Animal Science Journal* 2024;47:242–51. https://doi.org/10.5398/TASJ.2024.47.2.242

3. Hadi SN, Chung RH. Estimation of Demand for Beef Imports in Indonesia: An Autoregressive Distributed Lag (ARDL) Approach. *Agriculture (Switzerland)* 2022;12. https://doi.org/10.3390/agriculture12081212

4. Kousiga T, Vadivu RS. Hierarchical clustering algorithms in data mining. *Int J Sci Dev Res.* 2019;4:1–3.

5. Sokhonn L, Park YS, Lee MK. Hierarchical clustering via single and complete linkage using fully homomorphic encryption. *Sensors.* 2024;24. https://doi.org/10.3390/s24154826

6. Zhou S, Liu F, Song W. Estimating the optimal number of clusters via internal validity index. *Neural Process Lett.* 2021;53:1013–34. https://doi.org/10.1007/s11063-021-10427-8

7. Thrun MC, Stier Q. Fundamental clustering algorithms suite. *SoftwareX.* 2021;13:100642. https://doi. org/10.1016/j.softx.2020.100642

8. Radovanovic A, Li J, Milanovic JV, Milosavljevic N, Storchi R. Application of agglomerative hierarchical clustering for clustering of time series data. *IEEE PES Innov Smart Grid Technol Conf Eur.* 2020;2020-Oct:640–4. https://doi.org/10.1109/ISGT-Europe47291.2020.9248759

9. Zhang Y, Shi Q, Zhu J, Peng J, Li H. Time series clustering with topological and geometric mixed distance. *Mathematics.* 2021;9. https://doi.org/10.3390/math9091046

10. Wei WWS. *Time series analysis: univariate and multivariate methods.* 2nd ed. New York: Pearson Education; 2006.

11. Rahman NHA, Lee MH, Suhartono, Latif MT. Artificial neural networks and fuzzy time series forecasting: an application to air quality. *Qual Quant.* 2015;49:2633–47. https://doi.org/10.1007/s11135-014-0132-6

12. Monika P, Ruchjana BN, Abdullah AS, Budiarto R. Integration of GSTARIMA model with heteroskedastic error and kriging for climate forecasting: a systematic review. *Appl Math Inf Sci.* 2024;18:551–67. https://doi.org/10.18576/amis/180307

13. Ahmed SE, Aydin D, Yilmaz E. Semiparametric time-series model using local polynomial: an application on the effects of financial risk factors on crop yield. *J Risk Financ Manag.* 2022;15. https://doi.org/10.3390/ jrfm15030141

14. Montgomery DC, Friedman DJ. Prediction using regression models with multicollinear predictor variables. *IIE Trans.* 1993;25:73–85. https://doi.org/10.1080/07408179308964293

15. Arashi M, Norouzirad M, Roozbeh M, Mamode Khan N. A high-dimensional counterpart for the ridge estimator in multicollinear situations. *Mathematics.* 2021;9. https://doi.org/10.3390/math9233057

16. Ismail NH, Ahmad N, Mohamed NA, Tahar MR. Analysis of geomagnetic ap index on worldwide earthquake occurrence using principal component analysis and hierarchical cluster analysis. *Sains Malays.* 2021;50:1157–64. https://doi.org/10.17576/jsm-2021-5004-25

17. Maugeri A, Barchitta M, Favara G, La Mastra C, La Rosa MC, Magnano San Lio R, et al. Application of clustering on principal components for nutritional epidemiology: a workflow to derive dietary patterns. *Nutrients.* 2023;15. https://doi.org/10.3390/nu15010195

18. Wang T, Jing Z, Zhang S, Qiu C. Utilizing principal component analysis and hierarchical clustering to develop driving cycles: a case study in Zhenjiang. *Sustainability (Switzerland).* 2023;15. https://doi.org/10.3390/ su15064845

19. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 2010;11. https://doi.org/10.1186/1471-2156-11-94

20. Chan JY Le, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong ZW, et al. Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics.* 2022;10. https://doi. org/10.3390/math10081283

21. Penkova TG. Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety. *Procedia Comput Sci.* 2017;112:99–108. https://doi.org/10.1016/j.procs.2017.08.179

22. Contreras JM, Molina Portillo E, Fernández Luna JM. Evaluation of hierarchical clustering methodologies for identifying patterns in timeout requests in EuroLeague basketball. *Mathematics.* 2025;13. https://doi.org/10.3390/math13152414

23. Saraçli S, Doğan N, Doğan I. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J Inequal Appl.* 2013;2013. https://doi.org/10.1186/1029-242X-2013-203

24. Xie WB, Lee YL, Wang C, Chen DB, Zhou T. Hierarchical clustering supported by reciprocal nearest neighbors. *Inf Sci.* 2020;527:279–92. https://doi.org/10.1016/j.ins.2020.04.016

25. Vijaya, Aayushi S, Bateja R. A review on hierarchical clustering algorithms. *J Eng Appl Sci.* 2017;12:7501–7.

26. Lu G, Tian Z, Chen P, Liang Z, Zeng X, Zhao Y, et al. Comprehensive morphological and molecular insights into drought tolerance variation at germination stage in *Brassica napus* accessions. *Plants.* 2024;13. https://doi.org/10.3390/plants13233296

27. Dani ATR, Wahyuningsih S, Rizki NA. Penerapan hierarchical clustering metode agglomerative pada data runtun waktu. *Jambura J Math.* 2019;1:64–78. https://doi.org/10.34312/jjom.v1i2.2354

28. Herden T, Bunzel S. Archetypes of supply chain analytics initiatives—an exploratory study. *Logistics.* 2018;2:10. https://doi.org/10.3390/logistics2020010

29. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987;20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

30. Shutaywi M, Kachouie NN. Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy.* 2021;23. https://doi.org/10.3390/e23060759

31. Babii A, Ghysels E, Striaukas J. Machine learning time series regressions with an application to nowcasting. *J Bus Econ Stat.* 2022;40:1094–106. https://doi.org/10.1080/07350015.2021.1899933

32. Suhartono, Ashari DE, Prastyo DD, Kuswanto H, Lee MH. Deep neural network for forecasting inflow and outflow in Indonesia. *Sains Malays.* 2019;48:1787–98. https://doi.org/10.17576/jsm-2019-4808-26

33. Xu W, Peng H, Zeng X, Zhou F, Tian X, Peng X. A hybrid modelling method for time series forecasting based on a linear regression model and deep learning. *Appl Intell.* 2019;49:3002–15. https://doi.org/10.1007/s10489-019-01426-3

34. Budiantara IN, Ratnasari V, Ratna M, Zain I. The combination of spline and kernel estimator for nonparametric regression and its properties. *Appl Math Sci.* 2015;9:6083–94. https://doi.org/10.12988/ams.2015.58517

35. Ratnasari V, Budiantara N, Dani ATR. Nonparametric regression mixed estimators of truncated spline and Gaussian kernel based on CV, GCV, and UBR methods. *Int J Adv Sci Eng Inf Technol.* 2021;11:2400–6. https://doi.org/10.18517/ijaseit.11.6.14464

36. Draper NR, Smith H. Applied regression analysis. 3rd ed. Canada: John Wiley & Sons; 1998.

37. Montgomery DC, Peck EA, Vining GG. *Introduction to linear regression analysis.* 5th ed. Canada: John Wiley & Sons; 2012.

38. Fitriyah AT, Chamidah N, Saifudin T. Prediction of paddy production in Indonesia using semiparametric time series regression least square spline estimator. *Data Metadata.* 2025;4:1–12. https://doi.org/10.56294/dm2025527

39. Fibriyani V, Chamidah N, Saifudin T. Estimating time series semiparametric regression model using local polynomial estimator for predicting inflation rate in Indonesia. *J King Saud Univ Sci.* 2024;36. https://doi.org/10.1016/j.jksus.2024.103549

40. Alabdulrazzaq H, Alenezi MN, Rawajfih Y, Alghannam BA, Al-Hassan AA, Al-Anzi FS. On the accuracy of ARIMA-based prediction of COVID-19 spread. *Results Phys.* 2021;27:104509. https://doi.org/10.1016/j.rinp.2021.104509

41. Bank Indonesia. Strategic Food Price Information Center (PIHPS) [Internet]. n.d. Available from: https://www.bi.go.id/hargapangan (accessed January 16, 2023 and September 8, 2023).

42. Komalawati K, Malik A, Jatuningtyas RK, Wulanjari ME. Price transmission on the Indonesian beef market. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing Ltd; 2021. https://doi.org/10.1088/1755-1315/788/1/012209

43. Wibowo RP, Pebriyani D, Indrawati T, Khaliqi M. How Beef, Chicken, and Egg Demand Elasticities Vary with Income in Indonesia. Jurnal AGRISEP: Kajian Masalah Sosial Ekonomi Pertanian dan Agribisnis. 2025 Mar 14;361–84. https://doi.org/10.31186/jagrisep.24.01.361-384.

44. Fatimah SN, Zainuddin AF, Mardiana N, Mukhaiyar U. Beef Price Forecasting Based on Temporal, Spatial and Space-Time Parameter Indices. Barekeng 2025;19:1805–24. https://doi.org/10.30598/barekengvol19iss3pp1805-1824

45. Guo J, Tanaka T. The effectiveness of self-sufficiency policy: International price transmissions in beef markets. Sustainability (Switzerland) 2020;12. https://doi.org/10.3390/su12156073