RESEARCH ARTICLE • OPEN ACCESS

Comparative Analysis of Hierarchical Cluster Methods in Inflationary Cities in Indonesia Based on Sectoral Inflation Patterns

Husna Afanyn Khoirunissa et al.

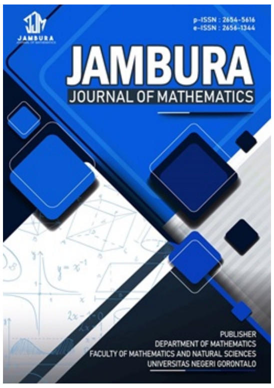
JAMBURA
JOURNAL OF MATHEMATICS

Volume 8, Issue 1, Pages 45–51, February 2026

Received 1 November 2025, Revised 20 January 2026, Accepted 30 January 2026, Published 2 February 2026

To Cite this Article : H. A. Khoirunissa et al.,"Comparative Analysis of Hierarchical Cluster Methods in Inflationary Cities in Indonesia Based on Sectoral Inflation Patterns", *Jambura J. Math*, vol. 8, no. 1, pp. 45–51, 2026, *https://doi.org/10.37905/jjom.v8i1.35105*

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS



Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
Journal Abbreviation	:	Jambura J. Math.
Frequency	:	Biannual (February and August)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/jjom
Online ISSN	:	2656-1344
Editor-in-Chief	:	Hasan S. Panigoro
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
Google Scholar ID	:	iWLjgaUAAAAJ
Email	:	info.jjom@ung.ac.id

Jambura Journal of Mathematics, Volume 8, Issue 1, Pages 45–51, February 2026 https://doi.org/10.37905/jjom.v8i1.35105

Research Article

Check for updates

Comparative Analysis of Hierarchical Cluster Methods in Inflationary Cities in Indonesia Based on Sectoral Inflation Patterns

Husna Afanyn Khoirunissa[1,*], Nur Rezky Safitriani[2], Erlyne Nadhilah Widyaningrum[3], Rizka Amalia Putri[4], Morina A. Fathan[2], Nabilla Rida Tri Nisa[5]

[1]*Statistics Study Program, Universitas Sebelas Maret, Surakarta 57126, Indonesia*
[2]*Statistics Study Program, Universitas Tadulako, Palu 94148, Indonesia*
[3]*Statistics Study Program, Universitas Mulawarman, Samarinda 75242, Indonesia*
[4]*Statistics Study Program, Universitas Riau, Pekanbaru 28293, Indonesia*
[5]*Multimedia Engineering Technology Study Program, Politeknik Negeri Media Kreatif, Jakarta Selatan 12630, Indonesia*

ARTICLE HISTORY
Received 1 November 2025
Revised 20 January 2026
Accepted 30 January 2026
Published 2 February 2026

KEYWORDS
Comparative Analysis
Hierarchical Cluster
Regional Inflation in Indonesia
Sectoral Inflation
Clustering Validation

ABSTRACT. *This study aims to assess the performance of single linkage, complete linkage, and average linkage hierarchical clustering algorithms in grouping cities used as inflation benchmarks in Indonesia into clusters based on sectoral inflation patterns. The data utilized are 150 regencies/cities divided into 11 sectors that drive inflation, identified by BPS Indonesia. Prior to clustering, a distance analysis using Euclidean distances was conducted to measure similarity between regions. Evaluation of the optimal number of clusters was conducted by applying the stability measure approach (APN, AD, ADM, and FOM), which showed that creating five clusters produced the most stable results. The results of the analysis revealed that the single linkage approach had the lowest within-cluster to between-cluster standard deviation ratio compared to the other two approaches, which revealed a greater level of homogeneity between the clusters. From an economic perspective, this clustering pattern revealed impressive differences in sectoral inflation pressures between provinces, even between cities within a province. Consequently, the single linkage method is proposed as the optimal method for identifying spatial variations in sectoral inflation in Indonesia.*

1. Introduction

Inflation is one of the most important measures of the economic stability of a region. Its movement can affect individuals' purchasing power, welfare, and the course of government fiscal and monetary policies [1]. In Indonesia, inflation not only varies over time but also shows high spatial inequality among cities and across contributing sectors. Sectoral inequality mirrors structural and consumption differences among regions. Thus, knowledge of urban sectoral inflation trends is relevant as a foundation for the derivation of more targeted, regionally based inflation management policies [2]. However, the complexity arising from the multidimensional nature of sectoral inflation across cities makes it impossible to interpret the patterns underlying these trends through descriptive analysis. Therefore, a data-driven method is needed to classify cities that share similarities in their sectoral inflation, providing a systematic approach to understanding inflation patterns across regions. In fact, cluster analysis provides an appropriate method for handling the complexity arising from sectoral inflation among different cities.

Cluster analysis is a statistical technique often employed to group areas based on common economic and social features [3]. Different studies have employed this technique to cluster Indonesian regions using various indicators. A study on provincial clustering based on the Healthy Family Indicator using both non-hierarchical and hierarchical clustering techniques indicated differences in the resulting structures between the two methods [4]. Another comparison was also conducted between hierarchical and non-hierarchical clustering of cities in Java using the Human Development Index (HDI), where the two methods resulted in different cluster formations [5].

In regional economics, the clustering approach has also been used to identify regions with comparable inflation characteristics [6]. In Murjani et al. (2022), regencies and cities in South Kalimantan were categorized into groups in order to identify inflation reference cities [7]. Meanwhile, other studies have employed hierarchical cluster techniques [8, 9], demonstrating the versatility of the method across different aspects of spatial analysis. Hierarchical clustering has also been applied to distance-based clustering among provinces, which revealed comparable prevalence patterns [10]. In economics, cluster analysis of urban areas based on inflation has been conducted using various techniques. For instance, Sari et al. (2023) compared Ward and K-Means algorithms for clustering Indonesian cities based on inflation indicators [11], whereas K-Means was used to cluster inflation levels of 34 provincial capitals before and during the COVID-19 pandemic [12]. Furthermore, machine learning studies illustrate the application of clustering methods to determine association patterns between inflation and regional poverty levels in

*Corresponding Author.

Email : *husnafanynk@staff.uns.ac.id* (H. A. Khoirunissa)
Homepage : http://ejurnal.ung.ac.id/index.php/jjom/index / E-ISSN : 2656-1344

Indonesia [13–15].

These studies demonstrate that cluster analysis is a significant tool for examining inter-regional heterogeneity. However, most previous studies have been conducted at the provincial level or based on aggregate macroeconomic indicators. In the present study, hierarchical clustering techniques are applied at the city level using data from 150 inflation cities designated by the Central Statistics Agency (BPS). These inflation cities are officially selected urban areas that serve as reference locations for measuring consumer price changes and calculating the Consumer Price Index (CPI) in Indonesia. The designation of inflation cities is based on their economic significance, population size, market activity, and availability of price data, making them representative units for monitoring regional inflation dynamics. The analysis considers sectoral inflation patterns categorized into 11 sectors according to BPS classifications. The results are expected to reveal clearer structures and linkages of inflation among Indonesian cities, contribute to the literature on spatial inflation analysis, and provide insights for future regional inflation control policy formulation.

2. Methods

The data used in this analysis are monthly inflation data for September 2025. The inflation data consist of 11 sectors as research variables, as presented in Table 1. The objects of the study are 150 regencies/cities in Indonesia designated as inflation cities by the Central Statistics Agency (BPS) of Indonesia. The research data were obtained from the official website of the Central Statistics Agency (BPS), bps.go.id, which was updated on October 1, 2025.

Table 1. Data variables

No.	Variable	Sector
1.	X_1	Food, beverages, and tobacco
2.	X_2	Clothing and footwear
3.	X_3	Housing, water, electricity, and household fuels
4.	X_4	Supplies, household equipment, and routine maintenance
5.	X_5	Health
6.	X_6	Transportation
7.	X_7	Information, communication, and financial services
8.	X_8	Recreation, sports, and culture
9.	X_9	Education
10.	X_{10}	Food and beverage/restaurant provision
11.	X_{11}	Personal care and other services

The research steps began with initial data preprocessing, which involved analyzing statistical summaries and detecting outliers. Next, the Euclidean distance matrix was calculated, the optimal number of clusters was determined, and hierarchical cluster analysis was applied using the single linkage, complete linkage, and average linkage methods. After obtaining the members of each cluster, each method calculated the within-cluster standard deviation (S_w) and the between-cluster standard deviation (S_b) to determine the average value of the S_w/S_b ratio. The method with the smallest average ratio value was considered the best method for grouping districts/cities based on sectoral inflation values in Indonesia.

2.1. *Data Preprocessing*

In an analysis, data preprocessing is required by looking at the statistical summary of the data and detecting outliers. Outlier detection is carried out because the presence of outliers can interfere with the analysis [16, 17]. Causes of outliers include errors in data entry, testing errors, data processing errors, sampling errors, and natural data (not errors) [18]. Outliers caused by errors can be overcome by removing the outlier from the analysis, but if not caused by errors, outliers can be retained to maintain the characteristics of the data.

2.2. *Cluster Analysis*

Cluster analysis is an analysis that aims to place a set of objects into two or more groups based on the similarities of the objects based on various characteristics. This is done to obtain a proximity matrix, i.e., a symmetric matrix with the same number of objects in both rows and columns. One distance between objects that can represent a measure of similarity is the Euclidean distance [16]. The Euclidean distance is formulated in eq. (1).

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (v_{ik} - v_{jk})^2}. \tag{1}$$

2.3. *Hierarchical Method*

The hierarchical method is a clustering method based on the similarities found in the objects, where similar objects will be clustered together and form a tree with a clear hierarchy between objects [19]. The steps in the clustering algorithm using the agglomerative hierarchical algorithm are as follows:

a. In N clusters, each cluster contains a single entity and a symmetric matrix of distances (similarities) $D = \{d_{ij}\}$ using the Euclidean distance formula from eq. (1) with size $N \times N$. The Euclidean distance is then used as the dissimilarity measure, as it effectively captures the overall magnitude of differences in sectoral inflation rates among cities and is well suited for hierarchical clustering with continuous variables.
b. Find the distance matrix for the closest cluster pair.
c. Pair and label the clusters with new names, and reduce the columns and rows of the cluster pairs to one column and one row.
d. Repeat steps (b) and (c) for $(N-1)$ times.

In hierarchical cluster analysis, there are several methods, i.e., the single linkage method, complete linkage, and average linkage.

2.4. *Single Linkage*

Clustering using single linkage is based on the smallest distance between two objects that are made the first cluster, and so on. In this method, it is done by finding the smallest distance in $D = \{d_{ij}\}$ or min$\{d_{ij}\}$ and then combining them into corresponding objects [16].

2.5. *Complete Linkage*

Clustering using complete linkage is based on the furthest distance between two objects that are made into the first cluster, and so on. In this method, it is done by finding the furthest distance in $D = \{d_{ij}\}$ or max$\{d_{ij}\}$ and then combining them into corresponding objects [16].

2.6. Average Linkage

Clustering using average linkage is based on the average distance between objects that are made into the first cluster, and so on [16]. It starts by finding the shortest distance in $D = \{d_{ij}\}$ and combining the same objects with the formula presented in eq. (2).

$$\frac{1}{n_r n_s} \sum_i \sum_j d(x_i, x_j). \tag{2}$$

2.7. Best Method

According to Bunkers et al. [20], to compare the quality of the best clustering, the standard deviation value criteria are used, i.e., the standard deviation within the cluster (S_w) and the standard deviation between clusters (S_b). S_w and S_b are formulated in eq. (3) and eq. (4).

$$S_w = \frac{1}{c} \sum_{k=1}^{c} S_{kj}, \tag{3}$$

$$S_b = \left[\frac{1}{c-1} \sum_{k=1}^{c} \left(\bar{X}_{kj} - \bar{X} \right)^2 \right]^{1/2}. \tag{4}$$

The smaller the S_w value and the larger the S_b value, the better the method performs or has high homogeneity. The smallest average S_w/S_b ratio indicates the best clustering method [20].

3. Results and Discussion

This section presents the results of a hierarchical cluster analysis conducted on sectoral inflation data for cities in Indonesia. This analysis includes the cluster formation process, determining the optimal number of clusters, and interpreting the characteristics of each city group based on the sectors contributing to inflation.

3.1. Data Preprocessing

Prior to conducting the cluster analysis, a statistical summary of the data used in this study is presented in Table 2.

Table 2. Summary statistic

Variable	Mean	SD	Min	Median	Max
X_1	0.24630	1.212805	-3.51	0.29	3.62
X_2	-0.01715	0.394768	-3.64	0.00	1.45
X_3	0.07172	0.238469	-1.00	0.00	1.21
X_4	0.00258	0.244776	-1.31	0.01	0.78
X_5	0.12950	0.327799	-1.23	0.02	1.73
X_6	-0.05033	0.652654	-4.61	0.00	1.94
X_7	0.00066	0.117017	-0.73	0.01	0.60
X_8	0.03993	0.392385	-1.93	0.00	3.76
X_9	0.38130	2.077574	-5.85	0.00	21.50
X_{10}	0.09079	0.197742	0.00	0.00	1.39
X_{11}	1.14800	0.764616	-0.89	1.17	3.07

Based on Table 2, it can be seen that the Food, beverages, and tobacco (X_1), Transportation (X_6), Education (X_9), and Personal care and other services (X_{11}) sectors have a range between minimum and maximum values that is large enough to affect the mean and standard deviation values. The minimum value in the Food, beverages, and tobacco sector comes from Merauke Regency, while the maximum value comes from Padangsidimpuan City. In the transportation sector, the minimum and maximum values come from Jayawijaya Regency and Merauke Regency, respectively. Furthermore, in the education sector, the minimum and maximum values come from Tangerang City and Toli Toli Regency. Finally, in the personal care and other services sector, the minimum and maximum values come from Meulaboh Regency and Banyuwangi Regency.

An inflation rate that is too high or too low in a sector indicates an imbalance between supply and demand. This condition can indicate structural problems in the distribution of goods, production efficiency, or purchasing power in the related sector. In the transportation sector, the minimum and maximum values are occupied by two regions located on the same island, i.e., Papua Island. This indicates a very high inequality in transportation cost increases. In the education sector, inflation in Toli Toli Regency has a very extreme value and is an outlier. This means that prices for education components in the regency increased by 21.50% compared to the previous year (YoY). The outlier value was confirmed not to be caused by an error but rather due to the natural value of the data; therefore, the data for Toli Toli Regency was retained for further analysis.

3.2. Euclidean Distance Matrix

In cluster formation, a distance matrix between inflation cities was first constructed using data consisting of 150 districts/cities and 11 variables. The distance between districts/cities was calculated using the Euclidean distance formula in eq. (1). The results of the distance calculations are presented in Table 3.

The distance matrix calculation is used to determine the similarity between objects. The smaller the distance between two objects, the more similar they are.

3.3. Optimal Cluster Determination

Before determining cluster members using the three hierarchical analysis methods, the number of clusters was determined using the three methods: single linkage, complete linkage, and average linkage. In this study, the stability score formulated by Brock et al. [21] was used in cluster validation for hierarchical cluster analysis. Stability measures the consistency of the resulting clusters when one variable is removed at a time. The smaller the value, the more consistent the clustering results. The measured scores are:

a. *Average proportion of non-overlap (APN)*: measures the proportion of "average number of observations in the same cluster" between dendrograms created using complete data and data from which one variable has been omitted.
b. *Average distance (AD)*: the average distance between observations in the same cluster, when creating clusters with complete data and data from which one variable has been removed.
c. *Average distance between means (ADM)*: the average distance between cluster centers when creating clusters with complete data and data from which one variable has been removed.

Table 3. Distance matrix with Euclidean distance

City	Central Aceh	Meulaboh	Aceh Tamiang	⋯	Jayawijaya
Central Aceh	0.00000	3.348477	1.232842	⋯	4.659903
Meulaboh	3.348477	0.00000	3.558286	⋯	6.005015
Aceh Tamiang	1.232842	3.558286	0.00000	⋯	4.837685
⋮	⋮	⋮	⋮	⋱	⋮
Nabire	2.008283	4.337015	1.914445	⋯	4.717393
Jayawijaya	4.659903	6.005015	4.837685	⋯	0.00000

Table 4. Stability score for three hierarchical cluster analysis methods

Measurement	Single Linkage		Complete Linkage		Average Linkage	
	Min Score	Cluster	Min Score	Cluster	Min Score	Cluster
APN	0.0024	3	0.0421	3	0.0024	3
AD	2.3120	5	2.2673	5	2.2922	5
ADM	0.0421	3	0.1000	3	0.0426	3
FOM	0.6025	5	0.5987	5	0.6030	4

d. *Figure of merit (FOM)*: the average intra-cluster variance on deleted variables, where clustering is performed using the remaining undeleted columns.

This study will test the optimality score to determine the best number of clusters, considering 3 to 5 clusters. The results of the best score (minimum score) measurement and the most optimal number of clusters based on the stability method for the three hierarchical cluster analysis methods are presented in Table 4.

Based on the results in Table 4, it can be seen that the optimal number of clusters selected based on the APN and ADM values in the three methods is three clusters, while based on the AD and FOM values (except for the average linkage method) it is five clusters. In this analysis, the AD value is used, which provides an optimal number of five clusters. The selection based on the AD score is motivated by the consideration that this measure accounts for the average distance between observations within the same cluster [22]. Thus, the AD measure is more sensitive to internal variations within each cluster compared to other measures such as APN, ADM, or FOM. The optimal AD value for five clusters indicates that the formation of five groups can minimize the distance between cluster members more consistently, resulting in a more stable and well-separated cluster structure. From an economic perspective, the division into five clusters is considered more representative in depicting the diversity of sectoral inflation patterns across cities in Indonesia. Each cluster can represent regional characteristics with different inflationary pressures and distinct sectoral combinations. Therefore, the selection of five clusters is more appropriate for identifying regional variations in the context of the real economy.

3.4. Single Linkage Analysis

Clustering using the single linkage method is carried out by calculating the shortest distance between two objects based on the distance matrix that has been formed. Using the hierarchical agglomerative algorithm with the single linkage method, the clustering process is repeated (as presented by the dendrogram in Figure 1) until five clusters remain, with the clustering results shown in Table 5. The resulting dendrogram exhibits a chaining effect, characterized by the gradual merging of observations at relatively low distances, which is a typical feature of the single linkage approach.

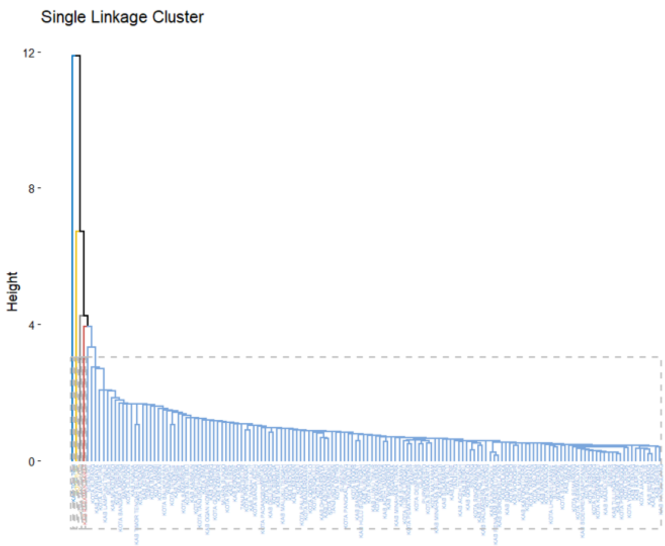


Figure 1. Single linkage dendogram

Cluster 1 has inflation rates for 11 sectors that are considered average, while Clusters 2, 3, 4, and 5 are filled by districts/cities with outlier inflation rates. Cluster 2 has high inflation rates in the Food, Beverages, and Tobacco; Recreation, Sports, and Culture; and Education sectors. Furthermore, Cluster 3 has high inflation rates in the Food and Beverage/Restaurant Provisions sector and the lowest inflation rate in the Education sector. Cluster 4 is characterized by the highest inflation rate in the Education sector and the lowest inflation rate in the Recreation, Sports, and Culture sector. Finally, Cluster 5 is almost identical to Cluster 4, with high inflation rates in the Education sector, although not as high as those in Cluster 4. Furthermore, in this cluster, inflation rates in the Transportation sector are very low.

Clusters 2, 3, 4, and 5 are each filled by only one regency/city representing a province. Cluster 2, West Pasaman Regency, is a regency in West Sumatra Province, where the province is also represented by several other cities, such as Dharmasraya

Table 5. Cluster member results of the single linkage method

Cluster	Cluster Members	Number of Members
Cluster 1	Central Aceh Regency, Meulaboh, Aceh Tamiang Regency, Banda Aceh City, Lhokseumawe City, …, Timika, Nabire Regency, Jayawijaya Regency	146
Cluster 2	West Pasaman Regency	1
Cluster 3	Tangerang City	1
Cluster 4	Toli Toli Regency	1
Cluster 5	Tual City	1

Regency, Padang City, and Bukittinggi City. Likewise, Tangerang City in Cluster 3 comes from Banten Province, which is represented by Pandeglang Regency, Lebak Regency, Cilegon City, and Serang City. Cluster 4, Toli Toli Regency, comes from Central Sulawesi Province, which is represented by Luwuk, Morowali Regency, and Palu City. Finally, Cluster 5, Tual City, comes from Maluku Province, which is also represented by Central Maluku Regency and Ambon City. All four clusters have other representatives from the same province, but these representatives are included in a different cluster (Cluster 1). This indicates that each city designated as a provincial representative for calculating inflation has different characteristics in each sector.

3.5. Complete Linkage Analysis

Clustering using the complete linkage method is performed by calculating the furthest distance between two objects based on the distance matrix that has been formed. Using the hierarchical agglomerative algorithm of the complete linkage method, the clustering process is repeated, as presented by the dendrogram in Figure 2, until five clusters remain, with the cluster results shown in Table 6. Compared to the single linkage method, the complete linkage approach forms more compact and well-separated clusters, since cluster merges occur at higher linkage distances.

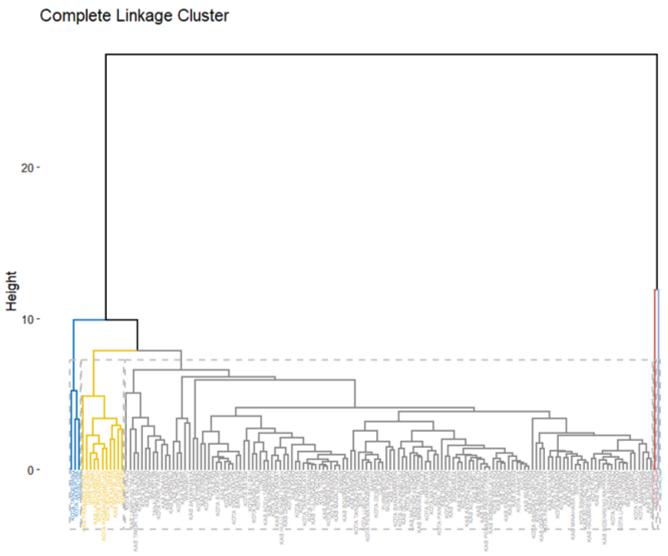


Figure 2. Complete linkage dendogram

Slightly different from the single linkage method, the results using the complete linkage method have more members in Cluster 2, with cluster member characteristics that are almost the same as the results in the single linkage method. Therefore, the average in the Food, Beverages, and Tobacco, Recreation, Sports, and Culture, and Education sectors is higher than the results in the single linkage method. Members in Cluster 3 increased by one district/city, i.e., Merauke Regency, with characteristics almost the same as Tangerang City, while Clusters 4 and 5 have the same members as the results in the single linkage method.

Table 6. Cluster member results of the single linkage method

Cluster	Cluster Members	Number of Members
Cluster 1	Central Aceh Regency, Meulaboh, Aceh Tamiang Regency, Banda Aceh City, Lhokseumawe City, …, Timika, Nabire Regency, Jayawijaya Regency.	108
Cluster 2	Labuhanbatu Regency, Deli Serdang Regency, Sibolga City, Pematangsiantar City, Medan City, …, North Minahasa Regency, Mamuju, Manokwari	38
Cluster 3	Tangerang City, Merauke	2
Cluster 4	Toli Toli Regency	1
Cluster 5	Tual City	1

3.6. Average Linkage Analysis

Clustering using the average linkage method begins by finding the shortest distance in $D = \{d_{ij}\}$ and combining the same objects with the formula in eq. (2) between two objects based on the distance matrix that has been formed. Using the hierarchical agglomerative algorithm of the average linkage method, the clustering process is repeated (presented by the dendogram in Figure 3) until 5 clusters remain, with the cluster results in Table 7. Figure 3 presents the dendrogram generated using the average linkage hierarchical clustering method, in which clusters are merged based on the average distance between all pairs of observations in two clusters.

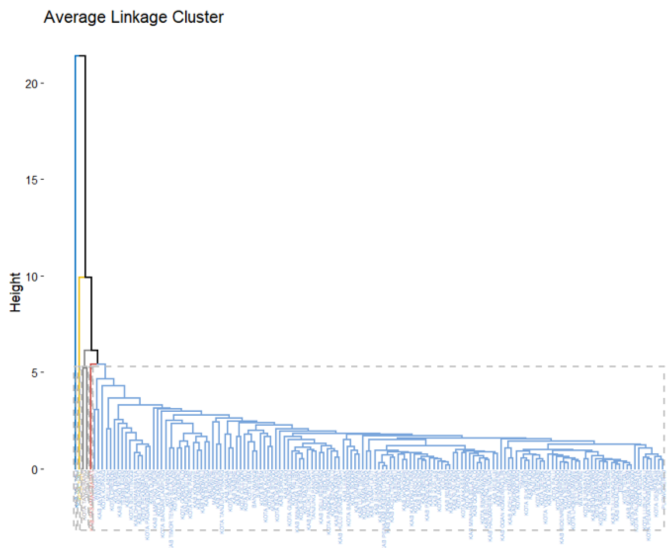


Figure 3. Average linkage dendogram

Table 7. Cluster member results of the average linkage method

Cluster	Cluster Members	Number of Members
Cluster 1	Central Aceh Regency, Meulaboh, Aceh Tamiang Regency, Banda Aceh City, Lhokseumawe City, …, Timika, Nabire Regency, Jayawijaya Regency.	145
Cluster 2	West Pasaman Regency	1
Cluster 3	Tangerang City, Merauke	2
Cluster 4	Toli Toli Regency	1
Cluster 5	Tual City	1

The results of this clustering are almost the same as the results of the single linkage method, with a slight difference in Merauke Regency. With the single linkage method, Merauke Regency is included in the members of Cluster 1, while with the average linkage method, it is included in the members of Cluster 3 (the same as the results with the complete linkage method). Therefore, the characteristics of members in Clusters 1, 2, 4, and 5 are the same as the characteristics produced by the single linkage method, while Cluster 3 is the same as the characteristics of Cluster 3 with the complete linkage method.

3.7. Comparison of Methods

A comparison of methods in cluster analysis was conducted to determine which agglomerative hierarchical method had the best performance in grouping inflation reference cities in Indonesia. In the comparison, the criteria of standard deviation within clusters (S_w) and standard deviation between clusters (S_b) were used, which were calculated using the formulas in eq. (3) and eq. (4), respectively.

Table 8. Standard deviation values

	Single Linkage	Complete Linkage	Average Linkage
S_w	0.0924	0.3152	0.2352
S_b	1.4860	1.2886	1.4631
S_w/S_b	0.0622	0.2446	0.1608

Table 8 shows that the standard deviation ratio of the single linkage method, 0.0622, is the smallest ratio among the hierarchical complete linkage and average linkage methods. Therefore, among the three methods used, the single linkage method has the best performance and high homogeneity within its clusters. Mapping using the single linkage method is presented in Figure 4.

As discussed in Section 3.4, the single linkage clustering results show substantial sectoral inflation pattern differences between Indonesian regencies/cities. Most areas are in clusters with comparable inflation rates, but a few are outliers because of extremely high inflation in a particular sector. This regional specificity means that price processes in any region are not necessarily the same, and even major cities within the same province will likely have their own patterns of inflation. This situation means that structural conditions and provincial-level consumption patterns also shape sectoral inflationary forces in any given region.

These results suggest that controlling inflation needs a more local and focused strategy. Provincial authorities need to

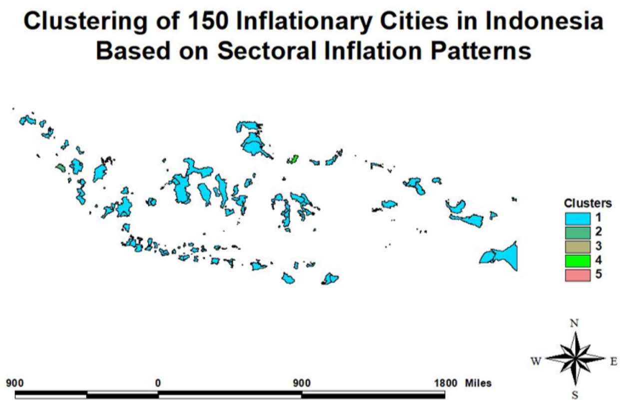


Figure 4. Mapping of clusters

enhance the monitoring of inflation to sectoral levels so that the policy may be effectively targeted. Second, the custom of considering one reference city needs to be re-evaluated, especially in the case of a heterogeneous region with regard to economic characteristics. Utilize the reference of city diversification and inter-regional cooperation by the Regional Inflation Control Team (TPID) as an excellent method for ensuring price stability, especially in food, education, and services, which are most susceptible to fluctuations.

4. Conclusion

The outcome of this research is hierarchical, with single linkage better at detecting differences among sectors between regions than the other two methods. The use of the shortest object-to-object distance in the clustering process makes it possible to detect fine differences among cities, and the shape of the resulting clusters becomes more elastic with regard to sectoral heterogeneity data. Conversely, the complete and average linkage algorithms produce more uniform cluster structures; however, they are less effective in distinguishing regions characterized by extreme or atypical features. Single linkage can thus be recommended as a superior algorithm to use to study clustering of economic data that is highly heterogeneous, like sectoral inflation in Indonesian cities.

These results suggest that inflation control has to be more context-specific and focused. Sectoral inflation tracking has to be bolstered by regional governments to allow more focused policies. In addition, having one reference city for provincial inflation calculation also needs to be addressed, especially where economic features are varied. Citing urban diversification and coordination among regions by the Regional Inflation Control Team (TPID) can be an ideal way of ensuring price stability, especially in the sectors with the highest tendencies to be affected by price instability, that is, food, education, and services.

This research contributes to the literature by offering an empirical comparison of hierarchical clustering techniques applied at the city level and using sectoral inflation data in Indonesia, a relatively underexplored region. By illustrating the enhanced sensitivity of single linkage for identifying sectoral heterogeneity in inflationary cities, this research provides methodological guidance for choosing appropriate clustering techniques when analyzing regional economies. Additionally, this research contributes to policy-oriented inflation research by illustrating the disadvantages of utilizing single reference cities at the provincial level and emphasizing the significance of sectoral and spatially differentiated inflation management techniques.

Author Contributions. **Husna Afanyn Khoirunissa**: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision. **Nur Rezky Safitriani**: Conceptualization, validation, supervision, editing. **Erlyne Nadhilah Widyaningrum**: Conceptualization, validation, supervision, editing. **Rizka Amalia Putri**: Conceptualization, validation, supervision, editing. **Morina A. Fathan**: Conceptualization, validation, supervision, editing. **Nabilla Rida Tri Nisa**: Conceptualization, validation, supervision, editing. All authors have reviewed and approved the final published version of this paper.

Acknowledgement. The authors would like to thank all those who contributed to this research and the preparation of the manuscript. We greatly appreciate the editors and reviewers for their input and support in improving this work.

Funding. This research did not receive external funding.

Conflict of interest. The authors declare no conflicts of interest related to this article.

Data availability. The data for this study are secondary data from the Central Statistics Agency. It is publicly accessible through https://www.bps.go.id.

References

[1] H. A. Khoirunissa, A. R. Wijaya, B. Isnaini, and K. Ferawati, "Analisis Faktor-Faktor Penyebab Inflasi di Indonesia Menggunakan Regresi Ridge, LASSO, dan Elastic-Net," *Indones. J. Appl. Stat.*, vol. 7, no. 2, p. 121, Jan. 2025, doi: 10.13057/ijas.v7i2.96921.

[2] F. Ferrante, S. Graves, and M. Iacoviello, "The inflationary effects of sectoral reallocation," *J. Monet. Econ.*, vol. 140, pp. S64–S81, Nov. 2023, doi: 10.1016/j.jmoneco.2023.03.003.

[3] A. Jaeger and D. Banks, "Cluster analysis: A modern statistical review," *WIREs Comput. Stat.*, vol. 15, no. 3, May 2023, doi: 10.1002/wics.1597.

[4] I. Nur and L. Fitriana, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Keluarga Sehat Menggunakan Metode Klaster Hirarki dan Non Hirarki," *J. Paradig. J. Multidisipliner Mhs. Pascasarj. Indones.*, vol. 2, no. 1, pp. 27–36, 2021. [Online]. Available: https://journal.ugm.ac.id/paradigma/article/view/66072

[5] A. R. Damayanti and A. W. Wijayanto, "Comparison of Hierarchical and Non-Hierarchical Methods in Clustering Cities in Java Island using the Human Development Index Indicators year 2018," *Eig. Math. J.*, pp. 8–17, Jun. 2021, doi: 10.29303/emj.v4i1.89.

[6] A. Setiawan, B. Susanto, and T. Mahatma, "Inflation data clustering of some cities in Indonesia," *J. Phys. Conf. Ser.*, vol. 855, p. 012046, Jun. 2017, doi: 10.1088/1742-6596/855/1/012046.

[7] A. Murjani, A. Pramila, and A. Rusyiana, "Analisis Klaster Kabupaten dan Kota di Kalimantan Selatan untuk Penentuan Kota Inflasi Acuan," *Ecoplan*, vol. 5, no. 1, pp. 53–63, Apr. 2022, doi: 10.20527/ecoplan.v5i1.429.

[8] S. Hidayatullah and A. Sofro, "Hierarchical Cluster Analysis Based on Waste Sources in Indonesia in 2022," *ComTech Comput. Math. Eng. Appl.*, vol. 15, no. 2, pp. 93–99, Nov. 2024, doi: 10.21512/comtech.v15i2.11088.

[9] M. F. F. Mardianto *et al.*, "Grouping of provinces in Indonesia based on community welfare level indicators using hierarchical cluster analysis," 2023, p. 080015, doi: 10.1063/5.0181024.

[10] S. Wulandari, "Clustering Indonesian Provinces on Prevalence of Stunting Toddlers Using Agglomerative Hierarchical Clustering," *Fakt. Exacta*, vol. 16, no. 2, Jul. 2023, doi: 10.30998/faktorexacta.v16i2.17186.

[11] L. P. Sari, A. Fanani, and A. H. Asyhar, "Analisis Perbandingan Pengelompokan Kota di Indonesia Berdasarkan Indikator Inflasi Tahun 2021 dengan Metode Ward dan K-Means," *J. Sains Mat. dan Stat.*, vol. 9, no. 2, p. 108, Aug. 2023, doi: 10.24014/jsms.v9i2.21100.

[12] N. Etrisia, M. F. Alexandi, and A. Asmara, "Klasifikasi Inflasi 34 Ibukota Provinsi di Indonesia Sebelum dan Saat COVID-19 Melalui Pengelompokan Wilayah dengan K-Means Clustering," *J. Ekon. Pembang.*, vol. 12, no. 2, pp. 120–133, Jul. 2023, doi: 10.23960/jep.v12i2.1597.

[13] R. Gustriansyah, J. Alie, A. Sanmorino, R. Heriansyah, and M. N. Megat Mohamed Noor, "Machine Learning for Clustering Regencies-Cities Based on Inflation and Poverty Rates in Indonesia," *Indones. J. Inf. Syst.*, vol. 5, no. 1, pp. 64–73, Aug. 2022, doi: 10.24002/ijis.v5i1.5682.

[14] J. Salomo and B. Siregar, "Cluster Analysis of Poverty Data in Cities/Districts in Indonesia Using K-Means Algorithm for the Years 2019–2022," 2025, pp. 487–495, doi: 10.1007/978-981-97-3859-5_37.

[15] S. Annas, B. Poerwanto, S. Sapriani, and M. F. S, "Implementation of K-Means Clustering on Poverty Indicators in Indonesia," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 257–266, Mar. 2022, doi: 10.30812/matrik.v21i2.1289.

[16] B. Simamora, *Analisis Multivariat Pemasaran*. Jakarta: Gramedia Pustaka Utama, 2005.

[17] A. Smiti, "A critical overview of outlier detection methods," *Comput. Sci. Rev.*, vol. 38, p. 100306, Nov. 2020, doi: 10.1016/j.cosrev.2020.100306.

[18] C. C. Aggarwal, "An Introduction to Outlier Analysis," in *Outlier Analysis*. Cham: Springer International Publishing, 2017, pp. 1–34, doi: 10.1007/978-3-319-47578-3_1.

[19] C. Essary, L. M. Fischer, and E. Irlbeck, "A Statistical Approach to Classification: A guide to hierarchical cluster analysis in agricultural communications research," *J. Appl. Commun.*, vol. 106, no. 3, Nov. 2022, doi: 10.4148/1051-0834.2431.

[20] M. J. Bunkers, J. R. Miller, and A. T. DeGaetano, "Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique," *J. Clim.*, vol. 9, no. 1, pp. 130–146, Jan. 1996, doi: 10.1175/1520-0442(1996)009<0130:DOCRIT>2.0.CO;2.

[21] G. Brock, V. Pihur, S. Datta, and S. Datta, "clValid: An R Package for Cluster Validation," *J. Stat. Softw.*, vol. 25, no. 4, 2008, doi: 10.18637/jss.v025.i04.

[22] X. Li, W. Liang, X. Zhang, S. Qing, and P.-C. Chang, "A cluster validity evaluation method for dynamically determining the near-optimal number of clusters," *Soft Comput.*, vol. 24, no. 12, pp. 9227–9241, Jun. 2020, doi: 10.1007/s00500-019-04449-7.