Jurnal Matematika dan Statistika serta Aplikasinya, Vol. 13, No. 2, 151-161, 2025
DOI:https://doi.org/10.24252/msa.v13i2.51488

Comparison of Cluster Analysis of Regency/City in South Sulawesi Province Based on Harvested Area and Rice Production Using the Average Linkage Method and Ward's Method

Nurul Jusmahilda Ismail[1], Try Azisah Nurman[2]*, Adiatma[3]
[1,2,3] *Mathematics Study Program, Universitas Islam Negeri Alauddin Makassar, Indonesia*
***Corresponding author**: try.azisah@uin-alauddin.ac.id

***Submission date**: 26 September 2024, **Revision**: 25 August 2025, **Accepted**: 21 November 2025

ABSTRACT
This study aims to further examine the two best methods, namely the Average Linkage Method and Ward's Method obtained from previous researchers by grouping 24 Regencies/Cities in South Sulawesi Province based on the variables of harvested area and rice production in 2023. In this study, Euclidean and Squared Euclidean distance measures were used. Furthermore, the minimum standard deviation value within the cluster (Sw), the maximum standard deviation between clusters (Sb), and the minimum ratio of Sw to Sb were observed to determine the most effective method. And also seen from the best Dunn Index value in determining the optimum number of clusters for each method. Based on the formation of the dendogram, the number of clusters tested was 2 to 6 clusters. The results of the study explain that of the two methods used, the method with the best performance is the Average Linkage method with the formation of an optimum cluster of 3 clusters, where cluster I has a low level of rice productivity. Cluster II has a high level of rice productivity. Cluster III has a fairly high (medium) level of rice productivity. And in the Ward's method, the optimum number of clusters formed is two clusters, with Cluster I having a low level of rice productivity and Cluster II having a high level of rice productivity.

KEYWORDS
Rice Harvest Area, Rice Production, Average Linkage Method, Ward's Method, Euclidean Distance and Squared Euclidean Distance.

1. INTRODUCTION

One of the regions in Indonesia that produces the most rice is South Sulawesi Province [1]. In addition to its crucial role in improving public welfare, the agricultural sector plays a crucial role in South Sulawesi's economic development, as demonstrated by its significant contribution to Gross Regional Domestic Product (PDRB). One priority in developing the agricultural sector, particularly food production in South Sulawesi, is improving the quality of agricultural products, particularly rice. This is evidenced by the high harvest area and rice production compared to other food crops.

Referring to official statistics news [1], The harvested area and rice production in South Sulawesi Province decreased in 2023. This was due to a decrease in rice productivity in the January-April 2023 and September-December 2023 sub-rounds [1]. According to ISIP news, this has impacted several regions, particularly in terms of productivity levels. Given the importance of the agricultural sector, particularly rice, in meeting the food needs of the South Sulawesi population, regional planning is needed to determine rice productivity in South Sulawesi Province by region (regency/city).

Various regional characteristics have a significant influence on the value of rice productivity produced in each Regency/City, both in terms of economics, environment and socio-culture [2]. By conducting a statistical analysis called cluster analysis, the characteristics of each region can be easily identified. This analysis should be conducted so that government policy programs can be more targeted and tailored to the characteristics of the clusters formed. The hierarchical method is a component of cluster analysis. In cluster analysis, the hierarchical method offers the advantage of faster data processing and time savings because the input data forms a dendogram or its own hierarchy, making it easier to interpret [3]. There are many methods for hierarchical cluster analysis and there is certainly one method that gives the best performance results in the cluster process.

This is in line with research conducted by Munthe (2019) which obtained results that the average linkage method is the method with the highest cophenetic correlation value [4]. This is also in line with research conducted by Riska (2022) that the advantage of the average linkage method is that it is able to group objects with small variations, is able to pay attention to the structure of the clusters formed and is more stable compared to the single linkage and complete linkage methods [5].

Natasya (2021) conducting a comparison of hierarchical methods evaluated by looking at the agglomerative coefficient value, where the Ward's method is able to provide the best cluster results [6]. This is also in line with research conducted by Riska (2022), which states that the advantage of the Ward's method is its ability to form clusters with low variance within the cluster, resulting in high homogeneity between objects. This method is stable, so it is often compared with other hierarchical methods [5]. Musarrafah (2019) also conducted research on two hierarchical cluster methods, namely the average linkage method and the Ward's method, which obtained the results that the average linkage method produced the best Dunn index value with a value of 0.55 compared to the Ward's method which produced a Dunn index value of 0.43, so the best method obtained was the average linkage method [7]. Imasdiani (2022) also conducted research comparing the Ward and average linkage methods, which found that the Ward's method had a lower average standard deviation ratio (R) than the average linkage method. Therefore, of the two methods tested, the Ward's method performed better [3].

Based on the description above, the author is interested in further researching the two best methods, the average linkage method and the Ward's method, obtained from previous researchers with different cases. This research was conducted by grouping regencies/cities in South Sulawesi Province based on harvested area and rice production variables.

2. LITERATURE REVIEW

2.1 Harvested Area and Rice Production

Rice production is calculated by multiplying the harvested area by productivity. The harvested area of rice in a rice field must be corrected by the equivalent number of galengan. Rice production is obtained by converting paddy production into rice using the conversion rate of unhusked rice and taking into account the proportion of unhusked rice/rice lost/wasted and used for non-edible purposes. Rice and rice production are calculated at the regency/city level. The harvested area of rice is calculated based on objective observations (objective measurements) using the KSA method developed by BPPT and BPS [1].

2.2 Rice Productivity

The Tiling Survey estimates are used to calculate rice productivity estimates. Since 2018, BPS has used KSA findings to determine the Tiling sample. Additionally, web-based data processing methods and software have been developed to check for outliers and improve the quality of the resulting data [1].

2.3 Multivariate Analysis and Cluster Analysis

Multivariate analysis is a technique for analyzing data using observations measured across multiple variables. Multivariate analysis is a statistical data analysis technique that takes into account correlations between variables and performs these analyses simultaneously. Multivariate data is defined as data consisting of more than one variable, which is used to measure a specific characteristic [3].

Cluster analysis is a multivariate technique that aims to group objects based on the characteristics they have in common [8]. Cluster analysis includes a technique for classifying or grouping objects or cases (respondents) into relatively homogeneous clusters [9]. Cluster analysis makes no assumptions about the number of clusters or their structure. Clustering is based on the similarity or distance between objects. Cluster analysis is divided into two areas: hierarchical and non-hierarchical clustering methods [10].

2.4 Distance Measure Between Data

The researchers compared the two methods because both are hierarchical clustering methods that aim to group objects based on similarity. And both have different principles in measuring distance and combining clusters, where the average linkage method uses the Euclidean distance measure because it calculates the average distance between all points in two different clusters, and the Euclidean distance measure inherently corresponds to the concept of the average. While the Ward's method uses the squared Euclidean distance because the main goal of this method is to minimize the increase in variance in each merged cluster, where the squared Euclidean distance gives more weight to larger differences in data values and is more suitable for the purpose of this method compared to the usual Euclidean [5].

1. Euclidean Distance

   Euclidean distance is the distance between objects on a straight line. For example, the inequality or distance measure between the i-th object and the j-th object can be symbolized by dij and for the k-th variable by $k = 1, \ldots, p$. The value of dij is obtained from the Euclidean distance calculation using the formula [11]:

$$d_{ij} = \sqrt{\sum_{k}^{p} (x_{ik} - x_{jk})^2} (2.1) \tag{1}$$

2. Squared Euclidean Distance

   The squared Euclidean distance is a variation of the Euclidean distance. The Euclidean distance has roots, while the squared Euclidean distance does not. The formula for this distance is [11]:

$$d_{ij} = \sum_{k}^{p} (x_{ik} - x_{jk})^2 \tag{2}$$

2.5 Hierarchical Clustering Method

The hierarchical methods are Agglomeration (merger) and Divisive (splitting) [12]. The agglomeration method begins with the assumption that each object is a cluster. Next, two adjacent objects are combined into one cluster. Then, a third object, taking into account the distance between them, will merge with an existing cluster or with another object to form a new cluster. It should be noted that in this method, each merger must be accompanied by a refinement of the distance matrix. A dendrogram is used to show the results of cluster analysis using this method. In contrast, divisive analysis begins with a large cluster containing all observations. Furthermore, non-identical observations are separated and grouped into smaller clusters. This process continues until each observation has its own cluster [13].

2.6 Average Linkage Method

The average linkage method is a clustering process based on the average distance or the average distance between objects [14]. This method is among the best of the hierarchical methods. However, compared to other hierarchical methods, this method requires the longest computation time [15]. The distance between two clusters can be thought of as the average distance between all members of one cluster and all members of the other cluster [16]. The goal of this method is to minimize the average distance between all pairs of observations and two connected clusters [17]. The first step that needs to be done is to first find the minimum distance $D = dij$ and combine it with the objects in question (such as U and V) to get the cluster (UV). The distance between other clusters W and (UV) is calculated using [17]:

$$d_{(UV)W} = average\ (d_{uw}, d_{vw}) \tag{3}$$

with,
$d_{(UV)W}$: distance between other cluster W with (UV)
$d_u w$: distance between cluster U with W
$d_v w$: distance between cluster U with V

2.7 Ward's Method

The ward's method is a clustering method based on information loss by grouping objects into clusters. This method is measured by the sum of the cluster mean squared deviations for each observation. The Sum of Squares Error (SSE) serves as the objective function. Two objects are combined if the objective function is the smallest among the available options. The SSE can be calculated when a cluster contains elements from multiple objects. A cluster with a single object is considered null. The Ward's method is calculated using a formula that follows the SSE formula [11]:

$$SSE = \sum_{i=1}^{N} \left(x_i - \bar{x}_{(m)}\right)^2 \tag{4}$$

with,
x_i: value of the i-th object with $i = 1,2,3,..,N$
$\bar{x}_{(m)}$: average value of the m-th cluster
N: number of objects in the formed cluster

2.8 Determining the Optimal Number of Clusters using the Dunn Index Value

Researchers also used the Dunn index value in this study to determine which method was most effective. The Dunn index validity function shows good assessment results for applications using various clustering methods. Cluster validity is determined by the fact that separate clusters typically have large distances between clusters and small distances within clusters. The highest value is the best way to find the best Dunn index. The following is the Dunn index formula [7]:

$$DI = \frac{{}_1 \leq a \leq b \leq K^{min}\ \{d\left(C_a, C_b\right)\}}{{}_1 \leq m \leq K^{max}\ \{d\left(C_m\right)\}} \tag{5}$$

with,
DI: Dunn index value
$d\left(c_a, c_b\right)$: distance between cluster a and b
$d(c_m)$: distance within cluster m

2.9 Selection of the Best Method with the Average Value of the Standard Deviation Ratio

In choosing the most effective method, the average standard deviation within a cluster (S_w) and the standard deviation between clusters (S_b) are used. The formula for S_w is as follows [17]:

$$S_W = \frac{1}{K} \sum_{m=1}^{K} S_m \tag{6}$$

with,
K : number of cluster formed
S_m : standard deviation of the m-th cluster

If given a cluster C_m then $m = 1,...,K$, each cluster has x_i members, where $i = 1,....,N$ and N are the number of members in each cluster, and $\bar{x}_{(m)}$ is the mean of cluster m. To find the m-th standard deviation value (S_m) use the formula [17]:

$$S_m = \sqrt{\frac{\sum_{=1}^{n} \left(x_{i(m)} - \bar{x}_{(m)}\right)^2}{N-1}} \tag{7}$$

If there is an average of variables in each cluster $m(\bar{x}_{(m)})$, then the components of each cluster will be different, and the standard deviation between clusters (S_b) is formulated as follows [11]:

$$S_b = \sqrt{\frac{1}{K-1} \sum_{m=1}^{K} \left(\bar{x}_{(m)} - \bar{X}^2\right)} \tag{8}$$

with,
K : number of cluster formed
$\bar{x}_{(m)}$: average of the m-th cluster
$\bar{X}$: average of all clusters

A good cluster method has a maximum S_b value and a minimum S_w value. The smaller the ratio of S_w to S_b, the better the method's performance because it has a high level of homogeneity within the cluster [11]. The ratio of S_w to S_b is calculated using the following formula:

$$R = \frac{S_W}{S_b} \tag{9}$$

3. METHODOLOGY

This research is an applied research. The data used in this study is secondary. The secondary data used comes from the Central Statistics Agency (BPS) of South Sulawesi Province in 2023 for the variables of harvested area and rice production. The tools used in this study are R Studio software and Microsoft Office Excel. Data standardization was the initial step in this research using z-scores. Descriptive analysis was then conducted to obtain an overview of the research data.Then, the calculation results of the average linkage and ward's methods are displayed in the form of a dendrogram. From this dendrogram, researchers form several clusters, which will then determine the optimum cluster using the Dunn index value. After that, the resulting clusters are interpreted. Finally, a comparison is made between the two methods used by calculating the standard deviation within the cluster (S_w) and the standard deviation between clusters (S_b) and the average standard deviation ratio for both methods.

4. RESULT & DISCUSSION

4.1 Descriptive Statistics

Descriptive statistics of each variable in the data used can be seen in the **Table 1**.

Table 1. Descriptive Statistics of Rice Harvest Area Variables in South Sulawesi Province

Statistics	Value
Minimum	962
Maximum	170330
Median	27063
Mean	40325
Standard Deviation	42347

Based on the **Table 1**, it is known that the lowest rice harvested area in South Sulawesi Province is 962 hectares, while the largest is 170,330 hectares. Descriptive statistics for rice production in South Sulawesi Province can be seen in the **Table 2**.

Table 2. Descriptive Statistics of Rice Production Variables in South Sulawesi Province

Statistics	Value
Minimum	3857
Maximum	847495
Median	120027
Mean	203183
Standard Deviation	212053

Based on the **Table 2**, it is known that the highest amount of rice production in South Sulawesi Province is 847,495 tons of GKG and the lowest is 3,857 tons of GKG.

4.2 Euclidean and Squared Euclidean Distance Calculations

The following is a summary of the matrix results from the Euclidean distance calculations for the average linkage method using the **Equation** (1):

Table 3. Euclidean Distance Calculation Results

	1	**2**	**3**	$\cdots$	**24**
1	0	1.385	0.270	$\cdots$	0.055
2	1.385	0	1.115	$\cdots$	1.330
3	0.270	1.115	0	$\cdots$	0.216
4	0.751	0.637	0.482	$\cdots$	0.697
5	0.838	0.550	0.568	$\cdots$	0.784
6	1.536	0.157	1.265	$\cdots$	1.481
7	0.689	0.696	0.419	$\cdots$	0.635
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
24	0.055	1.330	0.216	$\cdots$	0

Based on the results of the Euclidean distance calculation, the smallest distance between regencys/cities is 0.01530, namely the distance between object cluster 7 (Sinjai) and object 10 (Barru), which then the two regencys are combined into one cluster.

The following is a summary of the matrix results from the calculation of the squared Euclidean distance for the Ward's method using the **Equation** (2):

Table 4. Results of Squared Euclidean Distance Calculation

	1	**2**	**3**	$\cdots$	**24**
1	0	1.919	0.073	$\cdots$	0.003
2	1.919	0	1.243	$\cdots$	1.770
3	0.073	1.243	0	$\cdots$	0.047
4	0.564	0.406	0.232	$\cdots$	0.486
5	0.702	0.303	0.323	$\cdots$	0.614
6	2.358	0.025	1.601	$\cdots$	2.194
7	0.475	0.485	0.176	$\cdots$	0.403
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
24	0.003	1.770	0.047	$\cdots$	0

Based on the squared Euclidean distance calculation, the smallest distance between regencies/cities is 0.00023, which is the distance between clusters 7 (Sinjai) and 10 (Barru). These two regencies are then combined into one cluster. After that, the distance is recalculated to form a new cluster.

4.3 Average Linkage Method

The formation of new clusters is calculated using **Equation** (3) based on table **Table 3**. The distance between regencys/cities involving the formation of new clusters is object 7 (Sinjai) and object 10 (Barru). The calculation process is continued until a new cluster is created from all objects that have been combined during the agglomeration process. A dendrogram can be used to illustrate the cluster formation process. In this study, the number of clusters to be formed was not specified to determine the initial grouping of regencies/cities for harvested area and rice production in South Sulawesi. Therefore, based on the dendrogram, the researchers formed several clusters, and then determined the optimum cluster with the best clustering results. This dendrogram was created using the average linkage method, as shown in the following **Figure 1**.

Based on the **Figure 1**, several clusters can be obtained, which will then be compared to determine the optimum number of clusters. In this study, several clusters will be formed: 2, 3, 4, until 6 clusters.

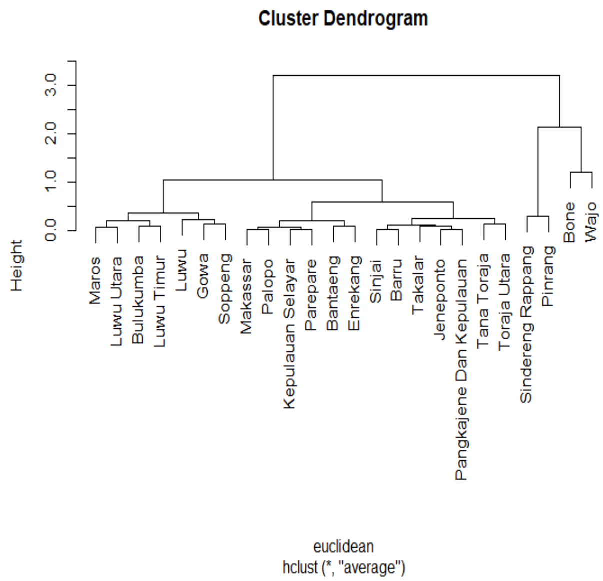

Figure 1. Dendogram of Average Linkage Method

4.4 Determining the Optimal Number of Clusters using the Dunn Index Value

Based on the analysis results, the cluster with the best cluster accuracy was determined. The cluster with the highest accuracy was then used as the initial cluster for harvested area and rice production in South Sulawesi. Cluster accuracy was determined using cluster validation with the Dunn index value, with the stipulation that the cluster with the highest Dunn index value was the best cluster to use. **Table 5** shows the accuracy results for each cluster formed.

Table 5. Dunn Index Value of Average Linkage Method

Number of Clusters	Dunn Index Value
2 Cluster	0.34338
3 Cluster	0.54407
4 Cluster	0.54406
5 Cluster	0.41234
6 Cluster	0.28704

Based on the **Table 5**, the Dunn index value for the optimum number of clusters of 3 is the highest, at 0.54407. Therefore, it can be concluded that the optimum number of clusters in regencies/cities in South Sulawesi based on harvested area and rice production using the average linkage method is 3 clusters. The cluster membership for the three clusters above is presented in the **Table 6**

Table 6. Cluster Membership of Average Linkage Method

Cluster	Regency/City
Cluster I (20 Members)	Kepulauan Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkajene dan Kepulauan, Barru, Soppeng, Enrekang, Luwu, Tana Toraja, Luwu Utara, Luwu Timur, Toraja Utara, Makassar, Pare-Pare, and Palopo
Cluster II (2 Members)	Bone and Wajo
Cluster III (2 Members)	Sidenreng Rappang and Pinrang

4.5 Interpretation of the Cluster Results that are Formed

Calculating the average value for each cluster will make it easier to identify its characteristics. **Table 7** shows the average values for the clusters formed:

Table 7. Average Value of Each Cluster Average Linkage Method

Cluster	Rice Harvest Area	Rice Production
Cluster I	24498.2	122604.9
Cluster II	155472.5	743155.5
Cluster III	83440.1	468990.5

1. In cluster I, the average harvested area and rice production are the lowest among the other clusters. Therefore, cluster I is categorized as a regency/city cluster with low rice productivity. This means that the regions in cluster I, including those with rice production potential, will experience a significant increase in harvested area and rice production by 2023 compared to other regions.

2. In cluster II, the average harvested area and rice production were the highest among the other clusters. Therefore, cluster II is categorized as a regency/city cluster with a high rice productivity level. This indicates that both variables influence rice productivity in South Sulawesi Province as a whole. This means that the regions in cluster II, including those with rice production potential, will experience a significant increase in harvested area and rice production by 2023 compared to other regions.

3. In cluster III, the average harvested area and rice production fall between clusters I and II. Therefore, cluster III is categorized as a mid-range regency/city cluster with a relatively high (moderate) level of rice productivity. This means that regions in cluster III, including those with rice production potential, will experience a significant increase in harvested area and rice production by 2023 compared to other regions.

4.6 Ward's Method

Agglomeration begins with the 2 objects closest to each other, the closest distance from the combination of 24 objects used. **Table 8** shows a summary matrix of the SSE values calculated between 2 objects.

Table 8. SSE Value Calculation Results between 2 Objects

	1	**2**	**3**	$\cdots$	**24**
1	0	0.959	0.037	$\cdots$	0.002
2	0.959	0	0.621	$\cdots$	0.885
3	0.037	0.621	0	$\cdots$	0.023
4	0.282	0.203	0.116	$\cdots$	0.243
5	0.351	0.151	0.161	$\cdots$	0.307
6	1.179	0.012	0.801	$\cdots$	1.097
7	0.238	0.242	0.088	$\cdots$	0.202
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
24	0.002	0.885	0.023	$\cdots$	0

Next, using the SSE values of the 2 objects formed above, the next step is to calculate the distance between a cluster and the new cluster that will be formed. The calculation process continues until the distances of all objects combined during the agglomeration process are calculated to form several clusters, resulting in a new cluster.

This dendrogram was created using the ward's method as shown in **Figure 2**: Based on the **Figure 2**, several clusters can be obtained, which will then be compared to determine the optimum number of clusters. In this study, several clusters will be formed: 2, 3, 4, until 6 clusters.

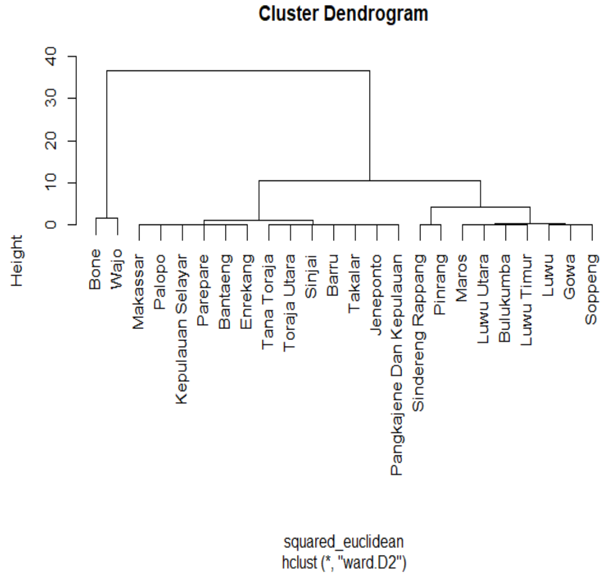

Figure 2. Dendogram of Ward's Method

4.7 Determining the Optimal Number of Clusters using the Dunn Index Value

Table 9 shows the accuracy results of each cluster formed:

Table 9. Dunn Index Value of Ward's Method

Number of Clusters	Index Dunn Value
2 Clusters	0.21307
3 Clusters	0.03573
4 Clusters	0.08545
5 Clusters	0.17003
6 Clusters	0.08239

Based on **Table 9**, the Dunn index value for the optimum number of clusters, 2, is the highest, at 0.21307. Therefore, it can be concluded that the optimal clusters in the regencies/cities in South Sulawesi based on harvested area and rice production using the Ward's method are 2 clusters. The cluster membership for the two clusters above is presented in the **Table 10**

Table 10. Cluster Membership of Ward's Method

Cluster	Regency/City
Cluster I (22 Members)	Kepulauan Selayar, Bantaeng, Bulukumba, Gowa, Maros, Soppeng, Sidenreng Rappang, Pinrang, Luwu, Luwu Utara, Luwu Timur, Jeneponto, Takalar, Sinjai, Pangkajene dan Kepulauan, Barru, Enrekang, Tana Toraja, Toraja Utara, Makassar, Pare-Pare, and Palopo
Cluster II (2 Members)	Bone and Wajo

4.8 Interpretation of the Cluster Results that are Formed

Table 11 shows the average value of the cluster formation that was formed :

1. In cluster I, the average harvested area and rice production were the lowest compared to cluster II. Therefore, cluster I is categorized as a Regency/City cluster with a low rice productivity level. This means that the achievements of

Table 11. Average Value of Each Cluster of Ward's Method

Cluster	Rice Harvest Area	Rice Production
Cluster I	29856.55	154094.50
Cluster II	155472.50	743155.50

regions in cluster I, including the potential for rice production in 2023, experienced a decrease in harvested area and low rice production compared to other regions. And has an SSE value of 14.993, which means the distance/similarity of characteristics between objects in cluster I is 14.993.

2. In cluster II, the average value of harvested area and rice production is the highest compared to cluster I. Thus, cluster II is categorized as a Regency/City cluster with a high level of rice productivity. This explains that both variables influence rice productivity in South Sulawesi Province as a whole. Which means that the achievements of the regions in cluster II, including the potential for rice production in 2023, experienced an increase in harvested area and high rice production compared to other regions. And has an SSE value of 0.760, which means the distance/similarity of characteristics between objects in cluster I is 0.760. Cluster II is the cluster with the best quality, because it has the smallest SSE value compared to the first cluster.

4.9 Determination of Comparison Results of Average Linkage Method and Ward's Method

The within-cluster standard deviation (S_w) and the between-cluster standard deviation (S_b) are two criteria used to measure the performance of a clustering method. A smaller within-cluster standard deviation and a larger between-cluster standard deviation indicate better performance of the method.

Table 12. Comparison of the Average Standard Deviation Ratio of the Two Methods

	Average Linkage Method	Ward's Method
Average Ratio Value	0.256	0.293

The average value of the standard deviation ratio (R) of the two methods is seen from the **Table 12**, obtained for both methods used, that the method with the best performance is the Average linkage method with an average value of the standard deviation ratio of 0.256 which is smaller than the Ward's method of 0.293. The smaller the value of the standard deviation ratio, indicates that the method used has better performance. The value of the standard deviation ratio obtained from both methods is at a standard deviation ratio value below 1 which means that both methods form a very homogeneous cluster annd based on the Dunn index values for the two methods above, the average linkage method has a higher value than the Ward's method. The higher the Dunn index value, the better the quality of the resulting clusters. Therefore, the method with the best performance is the average linkage method.

5. CONCLUSION

The study results show that of the two methods used, the average standard deviation ratio for the Average linkage method was smaller than the Ward's method. Therefore, the Average linkage method is the best clustering method.

REFERENCES

[1] BPS, "Berita resmi statistik no. 17/03/73/th. xxviii, 1 maret 2024: Luas panen dan produksi padi provinsi di sulawesi selatan 2023 (angka tetap)," 2024.

[2] N. Lestari, T. A. Nurman, and K. Nurfadilah, "Produktivitas Padi di Provinsi Sulawesi Selatan Menggunakan Metode Multidimensional Scaling (MDS)," *msa*, vol. 7, no. 1, p. 54, Aug. 2019.

[3] I. Imasdiani, I. Purnamasari, and F. D. T. Amijaya, "Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Average Linkage Dan Metode Ward," *J. Eksponensial*, vol. 13, no. 1, p. 9, Jun. 2022.

[4] A. D. Munthe, "PENERAPAN CLUSTERING TIME SERIES UNTUK MENGGEROMBOLKAN PROVINSI DI INDONESIA BERDASARKAN NILAI PRODUKSI PADI," *JLS*, vol. 2, no. 2, p. 11, May 2019.

[5] R. Ramadani and A. Salma, "PERBANDINGAN KINERJA METODE AVERAGE LINKAGE DAN WARD DALAM PENGELOMPOKAN TINGKAT KESEJAHTERAAN MASYARAKAT PROVINSI SUMATERA BARAT MENURUT KABUPATEN/KOTA TAHUN 2021," *UNPJoMath*, vol. 7, no. 3, p. 11, Sep. 2022.

[6] N. Afira and A. W. Wijayanto, "Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019," *Komputika*, vol. 10, no. 2, pp. 101–109, Sep. 2021.

[7] M. Paramadina, S. Sudarmin, and M. K. Aidid, "Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan)," *j. variansi*, vol. 1, no. 2, p. 22, Jul. 2019.

[8] A. Khoirunnisa', F. A. S. Wibowo, and K. Kismiantini, "Perbandingan Analisis Agglomerative Hierarchical Clustering Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat," *PSPMM*, vol. 7, Mar. 2023.

[9] B. S. A. Arif, A. Rusgiono, and A. Hoyyi, "PENGELOMPOKAN PROVINSI-PROVINSI DI INDONESIA MENGGUNAKAN METODE WARD (StudiKasus: Produksi Tanaman Pangan di Indonesia Tahun 2018)," *Gaussian*, vol. 9, no. 1, pp. 112–121, Feb. 2020.

[10] Hasmira, Wahidah Alwi, and Khalilah Nurfadilah, "Penentuan Cluster Hirarki Optimum dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan dengan Indikator Kemiskinan," *MSA*, vol. 11, no. 1, pp. 93–102, Jun. 2023.

[11] Musdalifah M. Ramly, S. Sudarmin, and B. Poerwanto, "Analisis Hierarchical Clustering Multiscale Bootstrap (Kasus: Indikator Kemiskinan di Provinsi Sulawesi Selatan Tahun 2020)," *j. variansi*, vol. 4, no. 3, pp. 142–152, Dec. 2022.

[12] A. N. Riadhoh, G. E. Puspita, I. Rafidah, and E. Widodo, "Pengelompokan Kabupaten/Kota Berdasarkan Produksi Tanaman Pangan Sumatera Utara Tahun 2020 Menggunakan Pengelompokan Hirarki Aglomeratif," *KUBIK*, vol. 6, no. 2, pp. 61–70, Apr. 2022.

[13] Y. Zahara, N. Afni Sinaga, R. Mahmuzah, and H. Hidayatsyah, "Analysis of Senior High Schools Quality in Banda Aceh Using Cluster and Correspondence Analysis," *PJPM*, vol. 16, no. 1, pp. 1–14, Jan. 2023.

[14] A. Zahra, A. A. R. Fernandes, A. Achmad, and R. Fernandes, "Prediction of late installment payment based on integration modeling cluster analysis average and ward linkage with survival analysis (case study: House ownership loan bank x customers)," 2022.

[15] R. Awaliah, "Analisis clustering untuk mengelompokkan tingkat kesejahteraan kabupaten berdasarkan sosial ekonomirumah tangga di wilayah provinsi sulawesi selatan." 2018.

[16] A. Saragih, D. C. Sukanda, and N. Eliyati, "CLUSTERIZATION OF REGION IN SOUTH SUMATERA BASED ON COVID-19 CASE DATA," *BAREKENG: J. Math. & App.*, vol. 17, no. 3, pp. 1257–1264, Sep. 2023.

[17] N. Ulinnuha and R. Veriani, "Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward," *InfoTekJar*, vol. 5, no. 1, 2020.