p-ISSN 2338 – 3216
e-ISSN 2528 – 1070

JURNAL STATISTIKA
UNIVERSITAS MUHAMMADIYAH SEMARANG
14(1) 2026: 01-21

Jurnal Statistika
Universitas Muhammadiyah Semarang

ASSESSING CLUSTER VALIDITY AND STABILITY OF HIERARCHICAL WARD'S LINKAGE AND NON-HIERARCHICAL K-MEANS ON THE EBGS INDEX OF REGENCIES/MUNICIPALITIES IN SOUTH SULAWESI

Elisabeth Evelin Karuna[1*], Mahrani[2], Atiqah Azza El Darman[3]
[1,2,3] Public Administration Study Program, Faculty of Social Sciences and Law, University of Makassar, Indonesia

***e-mail**: *elisabeth.evelin@unm.ac.id*

Article Info:
Received: October 10, 2025
Accepted: May 25, 2026
Available Online: June 2, 2026

Keywords:
Cluster; K-Means; Ward's Linkage; *Validity*; *Stability*; *EBGS*
.

Abstract: Cluster analysis is a statistical method used to group objects based on similar characteristics. In general, there are two main categories in cluster analysis, namely hierarchical methods (such as Ward's linkage) and non-hierarchical methods (such as K-Means). This study aims to compare the performance of these two methods in grouping the Electronic-Based Government System (EBGS) Index of districts/cities in South Sulawesi Province. The results of the analysis show that both methods produce identical validity index values, namely a Silhouette Coefficient of 0.67, a Davies-Bouldin Index (DBI) of 0.39, and a Calinski-Harabasz Index (CHI) of 83.02. These values indicate that the clusters formed have high internal compactness and clear separation between clusters. In addition, an Adjusted Rand Index (ARI) value of 1.00 indicates perfect agreement between Ward's linkage and K-Means, confirming the stability of the clustering structure. The analysis identifies three clusters representing Very Good (4 districts/cities), Good (10 districts/cities), and Fair (10 districts/cities) levels of EBGS implementation, showing that most regions fall into the Good and Fair categories. This clustering framework provides a basis for targeted policy interventions, enabling governments to priorities digital governance capacity building and infrastructure development in lower-performing regions while benchmarking best practices from high-performing regions. Overall, the results demonstrate that the EBGS clustering in South Sulawesi is statistically valid, stable, and representative of the underlying data structure.

1. INTRODUCTION

Cluster analysis is a statistical method that can be used to group certain characteristics. In general, cluster analysis is a statistical approach that includes various multivariate and quantitative methods aimed at grouping objects or observation units based on the degree of similarity between their characteristics [1]. Therefore, cluster analysis will certainly be very useful in this era of rapid information technology development, especially for researchers and practitioners to overcome the complexity of clustering various things in a more systematic and structured manner.

In general, there are two main categories in cluster analysis, namely non-hierarchical methods and hierarchical methods. Hierarchical cluster methods include several approaches

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

with varying characteristics, generally divided into two main types, namely agglomerative and divisive. Meanwhile, non-hierarchical cluster methods, such as K-Means, are one of the most commonly used approaches [2]. Based on research conducted by [3], the hierarchical clustering method shows superior performance compared to the non-hierarchical clustering method (K-Means) in grouping rainfall data. This approach is able to form clusters in which observations within one cluster have a higher degree of similarity than observations in other clusters. In line with these research findings, this study attempts to compare the performance of Ward's linkage (hierarchical) and K-Means (non-hierarchical) methods in a different context, namely in the grouping of the Electronic-Based Government System (EBGS) Index for districts/cities in South Sulawesi Province.

The Electronic-Based Government System (EBGS) is an innovative policy from the government to make it easier for the public to obtain optimal public services. However, in practice, the implementation of EBGS services in local governments often faces various obstacles that can lead to low implementation success rates. This situation certainly necessitates an evaluation of the level of EBGS implementation in each region through statistical measurement and grouping based on the EBGS Index [4]. Therefore, to support this evaluation, this study will use statistical analysis to compare the performance of Ward's linkage (hierarchical) and K-Means (non-hierarchical) methods in clustering the EBGS Index in South Sulawesi Province to assess the validity and stability of the resulting clusters. The main focus of this study is to assess the validity and stability of the clusters produced by both methods in order to obtain a more comprehensive understanding of the regional grouping structure based on the maturity level of EBGS implementation. Therefore, the objectives of this study are: (1) to form clusters of regencies or municipalities in South Sulawesi based on the EBGS index; (2) to compare Ward's linkage and K-Means clustering methods in terms of cluster validity using Silhouette Index (SI), Davies-Bouldin Index (DBI), and Calinski-Harabasz Index (CHI); and (3) to assess the stability of both clustering methods using the Adjusted Rand Index (ARI).

2. LITERATURE REVIEW

2.1 Cluster Analysis

Cluster analysis is a statistical approach that encompasses various multivariate and quantitative methods aimed at grouping objects or observation units based on the degree of similarity between their characteristics [1]. There are two main categories in cluster analysis, namely non-hierarchical methods and hierarchical methods. Hierarchical clustering methods include several approaches with varying characteristics. In general, these methods are divided into two main types, namely agglomerative and divisive. In the agglomerative approach, each object is initially treated as a separate cluster, then the most similar clusters are gradually combined until the desired number of clusters is reached. Conversely, the divisive approach starts with one large cluster that includes all objects, then gradually separates the objects into smaller clusters based on the degree of difference between objects. Meanwhile, non-hierarchical clustering methods, such as K-Means, are among the most commonly used approaches. This method works by dividing data into a number of clusters based on a predetermined k value. The clustering process is carried out iteratively to minimize the distance between each observation and the cluster center (centroid), resulting in data groups that have high homogeneity within clusters and heterogeneity between clusters [2].

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

In this study, both approaches were applied using Ward's Linkage as a representation of the hierarchical method and K-Means as a representation of the non-hierarchical method. The application of both approaches aims to evaluate and compare the validity and stability of the clustering results for the EBGS Index of districts/cities in South Sulawesi Province, so that a more comprehensive mapping of the maturity level of the implementation of electronic-based government systems in the region can be obtained.

2.2 Ward's Linkage Method

This method aims to produce clusters with the smallest possible internal variation. The clustering process is carried out based on an increase in the sum of squared error (SSE) value, where at each stage of merging, two clusters are selected that cause the smallest increase in SSE to then be merged into a new cluster [5]. In general, the SSE value can be calculated mathematically as follows [6].

$$I(G) = \sum_{x_i \in G} \|x_i - \bar{x}_G\|^2_{\mathbb{R}^p} \tag{1}$$

where $\bar{x}_G = n^{-1} \sum_{i=1}^{n} x_i$ denotes the centroid (center of gravity) of the group G. When two clusters G_u and G_v are merged, the resulting increase in total variance is expressed as follows:

$$\delta(G_u, G_v) := I(G_u \cup G_v) - I(G_u) - I(G_v) \tag{2}$$

The value $\delta(G_u, G_v)$ is called Ward's linkage, which represents the change in the sum of squared errors after two clusters are merged. The smaller the value of δ, the smaller the loss of homogeneity that occurs as a result of the merger. In another form, Ward's linkage can also be expressed as:

$$\delta(G_u, G_v) = \frac{|G_u||G_v|}{|G_u| + |G_v|} \|x_i - \bar{x}_G\|^2_{\mathbb{R}^p} \tag{3}$$

which shows that the merging of two clusters is based on the squared distance between the centers of the two clusters, taking into account the number of members in each cluster. The process in the Hierarchical Agglomerative Clustering (HAC) algorithm with the Ward's linkage method begins by considering each observation as a single cluster. Next, the two clusters with the smallest δ values are gradually merged until all data are combined into one large cluster. Thus, the Ward method produces clusters that are more compact, homogeneous, and have minimum dispersion within the cluster. The following is the Standard Hierarchical Agglomerative Clustering (HAC) algorithm [6]:

1. Initialization: Specify the initial partition
$$P_1 = \{\{x_1\}, \{x_2\}, \ldots, \{x_n\}\} \tag{4}$$
2. Iterative Merging: For each iteration $t = 1$ to $n - 1$: Calculate all linkage values between clusters in partition P_t; Merge the two clusters with the smallest linkage values to obtain the next partition P_{t+1}.
3. Termination: The procedure is repeated until all observations are merged into a single cluster, producing a sequence of partitions:
$$(P_1, P_2, \ldots, P_n) \tag{5}$$
This sequence represents all stages of cluster formation, from the initial partition (each object stands alone) to the final partition (all objects become one cluster).

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

Distance measurement is a key component in HAC. In this study, Manhattan distance is used as the dissimilarity measure, defined as [7]:

$$d_{manh}(a,b) = \sum_{k=1}^{n} |y_k - x_k| \tag{6}$$

where x_k and y_k re the values of each variable (or k-th dimension coordinate) in the two objects being compared.

2.3 K-Means Clustering

The K-Means method groups data into k separate clusters based on specific criteria using an iterative process that aims to maximize uniformity within clusters and differences between clusters [8]. The non-hierarchical K-Means method clusters all observations by utilizing specific starting points to determine cluster membership more precisely. This approach complements the advantages of hierarchical methods with the ability to refine clustering results through an iterative process that allows for changes in cluster membership until optimal convergence is achieved [3].

The K-Means model groups each observation into one of k cluster centers, where each center represents the average characteristics of that cluster. This process can be influenced by measurement errors in feature variables. Mathematically, this model can be expressed in matrix form as follows [9]:

$$\mathrm{X} = \mathrm{U}_k \mathrm{M}_k + \mathrm{E}_k \tag{7}$$

where U_k, M_k, and E_k represent the membership matrix, prototype matrix (cluster center), and error matrix, respectively. Mathematically, the K-Means clustering algorithm is presented as follows [10]:

1. Initialization: Randomly select a number of initial centers (centroids), namely $\mu_1, \mu_2, \ldots, \mu_k$.
2. Assignment (Cluster Determination): Each data point x will be placed in the cluster with the nearest centroid, using the formula:

$$S_i^{(t)} = \left\{ x : \left\| x - \mu_i^{(t)} \right\|^2 \leq \left\| x - \mu_j^{(t)} \right\|^2 \forall j, 1 \leq j \leq k \right\} \tag{8}$$

3. Update Centroid: Recalculate the center (mean) of each cluster based on the points that have entered the cluster:

$$\mu_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_i \in S_i} x_j \tag{9}$$

4. Check Convergence: Check whether the cluster centers have not changed or the number of iterations has reached the maximum limit. If:

$$\mu_i^{(t+1)} = \mu_i^{(t)} \ , \forall i \tag{10}$$

then the algorithm stops. If not, return to Step 2 and repeat the process.

One of the main challenges in K-Means is determining the optimal number of clusters (k). Choosing the right k is very important because too few clusters can obscure data patterns, while too many clusters can result in overfitting and complex interpretations. For this reason, there are several methods that are often used to determine the optimal k, including:

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

1. Silhoutte

In line with research conducted by [11], maximizing the Silhouette coefficient value as a measure of cluster validity is used to estimate the variability of the clustering results obtained. One of the main limitations of Silhouette is its computational cost. For center-based clustering algorithms such as K-Means and K-Medoids, a simple approach to calculating Silhouette can be done by using the distance to each cluster center or medoid [12].

2. Elbow

The Elbow method provides a simple approach to determining the most appropriate number of clusters in a data set. In this method, the initial number of clusters (n) is selected randomly. Then, the clustering algorithm in this study, K-Means, is run to divide the data into n clusters. Next, the Within-Cluster Sum of Squares (WSS) is calculated for each number of clusters used. The number of clusters n_p that produces the lowest WSS is considered the optimal number of clusters [13].

3. Gap Statistics

In the Gap Statistics (GS) approach, a random data distribution is first created as a zero reference, which is used as a comparative value to measure cluster compactness. The optimal number of clusters is determined at the point where the cluster compactness value has the largest difference compared to the reference curve. This point is then considered to be the most appropriate number of clusters [13].

After the optimal number of clusters has been obtained, the next step is to evaluate the validity and stability of the clusters in order to assess the extent to which the resulting clusters are representative and consistent with the data variation.

2.4 Cluster Validity and Stability

Evaluating cluster validity and stability is an important step in clustering analysis to ensure that the grouping results obtained are not only statistically optimal, but also structurally meaningful. Cluster validity is used to assess the extent to which the clusters formed truly reflect the natural separation between data, while cluster stability focuses on the consistency of clustering results when there are minor changes in the data or algorithm parameters. In this study, the Silhouette Coefficient, Davies-Bouldin Index (DBI), and Calinski-Harabasz Index (CHI) will be used to measure cluster validity, while the Adjusted Rand Index (ARI) will be used to measure cluster stability. These evaluation metrics were chosen because they provide a comprehensive overview of the quality of the data grouping results.

The Silhouette Index (SI) is used to assess the extent to which each data point is in the correct cluster and its proximit4y to other clusters. A higher SI value indicates that the data point is more consistent with the cluster it belongs to. The Davies-Bouldin Index (DBI) serves to measure the degree of separation between clusters, where a smaller DBI value indicates better cluster separation and a clearer structure. Meanwhile, the Calinski-Harabasz Index (CHI) is used to assess the degree of clarity and compactness of clusters. A high CHI value indicates that the clusters formed have strong differences between each other and a good level of uniformity within them. In this study, SI, DBI, and CHI values were calculated for each algorithm and variation in the number of clusters (k). In addition, the Adjusted Rand Index (ARI) was used to measure the degree of similarity between two different clustering results. ARI values range from -1 to 1, where 0 indicates random agreement and 1 indicates perfect agreement [14]. In this study, a high Calinski-Harabasz Index (CHI) value

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

is interpreted as evidence that the EBGS clusters among regencies/municipalities are clearly separated in terms of digital governance maturity, with strong homogeneity of EBGS performance within each cluster. Furthermore, an Adjusted Rand Index (ARI) value of 1.00 is particularly important as it indicates perfect agreement between Ward's linkage and K-Means clustering results, confirming that the identified clustering structure is highly stable and not dependent on the clustering algorithm used.

3. METHODOLOGY

This section will explain in detail the stages and approaches used in the research, starting from the data sources, research variables, data collection methods, to the analysis procedures applied. The explanation of each stage is arranged systematically in order to describe the overall research flow and ensure that the analysis process can be replicated properly.

3.1 Types and Sources of Data

The data used in this study is secondary data on the results of the evaluation of the local government's EBGS in South Sulawesi, obtained from the Decree of the Minister of Administrative and Bureaucratic Reform of the Republic of Indonesia concerning the Results of the Evaluation of the Electronic-Based Government System (EBGS) in Central Agencies and Local Governments in 2024.

3.2 Research Variables

The variable used in this study is the Electronic-Based Government System Index (EBGS) at the district/city level in South Sulawesi Province. The EBGS index value is used as the basis for the cluster analysis process to group districts/cities based on the level of EBGS implementation in the South Sulawesi region.

3.3 Analysis Procedure

The following are the stages of cluster analysis conducted using two approaches, namely Ward's Linkage hierarchical method and K-Means non-hierarchical method.

1. **Descriptive Statistics**
   This stage aims to provide an overview of the characteristics of the EBGS Index values for districts/cities in South Sulawesi Province. Descriptive statistics are used to examine the distribution, central tendency (mean and median), dispersion (quartiles and range), and shape of the data distribution (skewness and kurtosis). This analysis provides an initial assessment of the level and variability of digital governance maturity across regions.
2. **Data Visualization**
   The EBGS Index data were visualized using boxplots and thematic maps. The boxplot was employed to assess the distributional characteristics, median, interquartile range, and potential outliers. Based on the descriptive statistics, the distribution was slightly right-skewed (skewness = 0.56) but remained approximately symmetric, as indicated by the close values of the mean and median. The boxplot also allows identification of extreme observations that may influence statistical measures, particularly the mean. In addition, thematic maps were used to visualize the spatial distribution of EBGS Index values across districts/cities. This visualization provides preliminary insights into regional patterns and potential spatial clustering prior to formal clustering analysis.

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

3. **Cluster Analysis**
   Cluster analysis was performed using two approaches, namely Ward's Linkage hierarchical method and K-Means non-hierarchical method. Ward's Linkage method was used to determine the optimal number of clusters through dendrogram and distance between objects, while K-Means method was used to reinforce the grouping results based on the cluster centroids that had been formed. In this study, both Ward's linkage and K-Means clustering algorithms employ the Manhattan distance metric to measure dissimilarities among observations. The use of Manhattan distance is intended to provide robustness to outliers and to capture linear separations in the EBGS multidimensional space. By applying the same distance metric to both hierarchical and non-hierarchical methods, this study ensures that differences in clustering results are attributable to the clustering algorithms rather than the distance measure.
4. **Determination of the Optimal Number of Clusters**
   The optimal number of clusters was evaluated using three internal validation techniques: the Elbow method, Silhouette Index (SI), and Gap Statistic. The Elbow and Silhouette methods indicated that three clusters ($k = 3$) provided a meaningful partition of the data. Although the Gap Statistic suggested $k = 1$, reflecting overall homogeneity among regencies/cities in terms of the EBGS index, this study prioritised interpretability and practical usefulness for policy analysis. Therefore, $k = 3$ was selected as a conscious and transparent modelling choice to ensure meaningful differentiation among groups.
5. **Cluster Evaluation (Validity and Stability)**
   This stage aims to assess the quality of the clustering results. Cluster validity is measured using three indices, namely the Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index to assess the compactness and separation between clusters. Meanwhile, cluster stability is tested using the Adjusted Rand Index (ARI) to measure the consistency of results between the two clustering methods.

4. RESULTS AND DISCUSSION

This section presents the results of the analysis and discussion, including data description, visualization, cluster formation, and evaluation of the validity and stability of the clustering results for the EBGS Index of districts/cities in South Sulawesi Province.

4.1 Descriptive Statistics

Descriptive statistics are used as an initial step to describe or provide an overview of the data used. The data used in this study is the evaluation data of the Electronic-Based Government System (EBGS) of Regencies/Cities in South Sulawesi in 2024. The results of descriptive statistics are presented in Table 1**:**

Table 1. Statistics Descriptive

Min	Q1	Median	Mean	Q3	Max	Skewness	Kurtosis
2.31	2.72	3.02	3.02	3.19	4.02	0.56	2.99

The descriptive statistics in Table 1 show that the EBGS index for regencies/cities in South Sulawesi in 2024 ranges from 2.31 to 4.02. The average and median values are the same, namely 3.02, indicating that EBGS achievements are generally in the good category with a

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

fairly stable level of equity. Most regions have values in the range of 2.72 to 3.19 (IQR), which shows that the distribution of scores is relatively concentrated around the median value. This condition indicates that EBGS implementation in most regions tends to be uniform, with no significant differences between regions.

The data distribution is slightly right-skewed (skewness = 0.56), but overall the distribution remains close to symmetric, as indicated by the identical mean and median values. The kurtosis value of 2.99 suggests a distribution close to normal, indicating no extreme tail behaviour. Thus, although in general EBGS achievements have been consistent in the good category, there are disparities between regencies/cities that are important to note, especially for regions with below-average achievements so that they can catch up in the implementation of e-government. In addition to looking at data characteristics using descriptive statistics, we will further examine data characteristics using data visualization in the form of thematic maps and boxplots.

4.2 Data Visualization

After conducting descriptive statistical analysis to obtain an initial overview of the data characteristics, the next step is to present the data in the form of visualizations. Data visualization serves to clarify patterns, trends, and differences that may not be immediately apparent from the statistical figures. In this study, the visualizations are presented through thematic maps and boxplots.

a. Thematic Map

Thematic maps are used to illustrate the distribution of EBGS achievements based on regencies/cities in South Sulawesi.

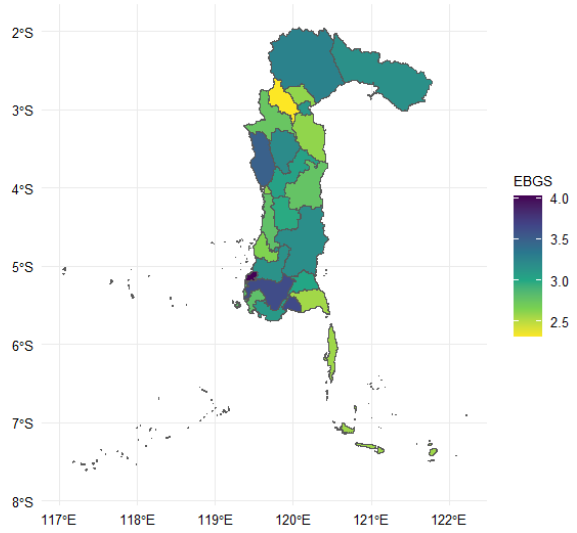


Fig 1. Thematic Map of EBGS South Sulawesi

Based on Fig 1 Thematic Map of EBGS Achievement in Regencies/Cities in South Sulawesi in 2024, it can be seen that most of the regions are in the good category, as indicated by the dominance of green to blue colors on the map. This confirms that the implementation of e-government is relatively evenly distributed across most regions. However, there are several regencies/cities that are shown in darker colors closer to dark blue, indicating higher achievements reaching the very good category.

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

Conversely, there are also one or two areas with bright yellow colors, which indicate that EBGS achievements are relatively lower than other regions and are still in the adequate category. This spatial pattern shows disparities between regions, where regions with higher achievements can be used as references or models for other regions, while regions with lower achievements need more attention in efforts to improve electronic-based governance at the regional level. Thus, the thematic map not only shows the spatial pattern of EBGS achievement, but also provides an initial picture of the possibility of extreme values, which will then be clarified through Boxplot analysis in the following section.

b. Boxplot

Boxplots are used to show the distribution of data, median values, and potential outliers from the analyzed EBGS scores.

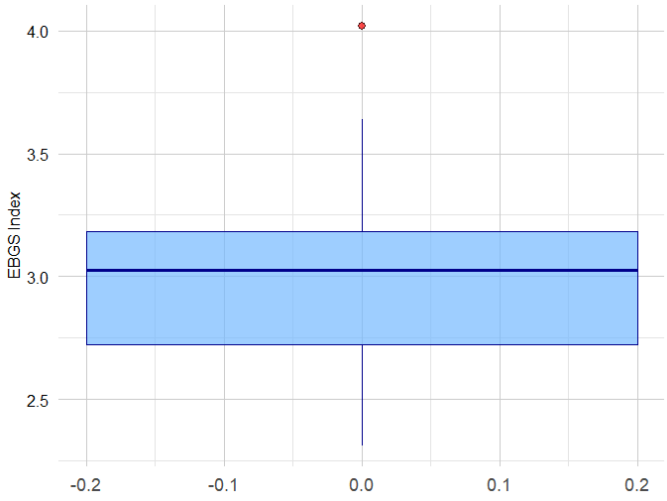


Fig 2. Boxplot EBGS

The boxplot in Fig 2 represents the distribution of the EBGS Index, which is generally concentrated in the range of 2.5 to 3.7, with the median close to 3.0. The relatively symmetrical position of the median within the interquartile range indicates that the data distribution is relatively balanced, with a slight right-skewness, consistent with the descriptive statistics. However, one data point with a value close to 4.0 was identified and classified as an outlier. The presence of this outlier reflects an observation with an EBGS Index that is substantially higher than the general pattern of the other data. Methodologically, the existence of outliers has important implications because it has the potential to affect descriptive parameters, particularly the mean value, which can shift due to the presence of these extreme values. Therefore, further identification and study of outliers is necessary to ensure that the interpretation of the analysis results remains accurate, representative, and does not cause analytical bias.

When comparing the thematic map in Fig 1 and the boxplot in Fig 2, there is consistency in the distribution pattern of EBGS achievements in South Sulawesi. The majority of regions have relatively uniform achievements, falling within the medium range of 2.5-3.5, as illustrated by the dominant colors on the map. Meanwhile, areas with higher scores, which appear as outliers in the boxplot, stand out in the thematic map with more contrasting colors. This reinforces the interpretation that most areas are at a balanced level of achievement, with one area deviating positively and warranting special attention in further analysis. Therefore, a statistical method is needed to strengthen the statistical interpretation

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

of the EBGS achievement level in South Sulawesi. The statistical method to be used is cluster analysis, which aims to group regencies/cities in South Sulawesi based on the Electronic-Based Government System (EBGS) service achievement index.

4.3 Cluster Analysis

In this study, cluster analysis was used to group districts/cities in South Sulawesi based on the level of achievement of the Electronic-Based Government System (EBGS). This analysis aims to identify patterns of similarity between regions so that areas with similar levels of achievement will be grouped together. Through this approach, a more comprehensive picture of the variation in EBGS achievement in each region can be obtained, enabling the formulation of more targeted improvement strategies according to the characteristics of each group. To obtain optimal clustering results, two cluster analysis methods were used, namely Ward's method and K-Means clustering.

a. Ward's Method

Ward's method is an agglomerative clustering technique based on the classical sum of squares principle. This approach aims to produce clusters with a high degree of internal uniformity by minimizing dispersion or variance within groups at each stage of the gradual (binary) merging process [15]. As is known, Ward's method is an agglomerative clustering technique based on the distance between the observations being analyzed. Based on the results in 4.2 (b), it is concluded that one data point with a value close to 4.0 was identified and classified as an outlier. Therefore, the appropriate distance to use when there are outliers is the Manhattan distance. This is in line with research conducted by [16], which concluded that outliers can be detected based on the MDBA (Manhattan Distance Based Algorithm) algorithm or can be said to be robust against outliers. The following presents the output from the Manhattan distance calculation:

Table 2. Manhattan Distance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0	0.85	0.46	1.08	0.44	0.01	0.54	1.07	0.38	0.52	1.02	1.04	0.48	0.49	1	0.15	0.6	0.63	0.66	0.84	0.89	1.33	0.87	0.38
2	0.85	0	0.39	0.23	0.41	0.84	0.31	0.22	1.23	0.33	0.17	0.19	0.37	0.36	0.15	0.7	0.25	0.22	0.19	0.01	0.04	0.48	0.02	0.47
3	0.46	0.39	0	0.62	0.02	0.45	0.08	0.61	0.84	0.06	0.56	0.58	0.02	0.03	0.54	0.31	0.14	0.17	0.2	0.38	0.43	0.87	0.41	0.08
4	1.08	0.23	0.62	0	0.64	1.07	0.54	0.01	1.46	0.56	0.06	0.04	0.6	0.59	0.08	0.93	0.48	0.45	0.42	0.24	0.19	0.25	0.21	0.7
5	0.44	0.41	0.02	0.64	0	0.43	0.1	0.63	0.82	0.08	0.58	0.6	0.04	0.05	0.56	0.29	0.16	0.19	0.22	0.4	0.45	0.89	0.43	0.06
6	0.01	0.84	0.45	1.07	0.43	0	0.53	1.06	0.39	0.51	1.01	1.03	0.47	0.48	0.99	0.14	0.59	0.62	0.65	0.83	0.88	1.32	0.86	0.37
7	0.54	0.31	0.08	0.54	0.1	0.53	0	0.53	0.92	0.02	0.48	0.5	0.06	0.05	0.46	0.39	0.06	0.09	0.12	0.3	0.35	0.79	0.33	0.16
8	1.07	0.22	0.61	0.01	0.63	1.06	0.53	0	1.45	0.55	0.05	0.03	0.59	0.58	0.07	0.92	0.47	0.44	0.41	0.23	0.18	0.26	0.2	0.69
9	0.38	1.23	0.84	1.46	0.82	0.39	0.92	1.45	0	0.90	1.4	1.42	0.86	0.87	1.38	0.53	0.98	1.01	1.04	1.22	1.27	1.71	1.25	0.76
10	0.52	0.33	0.06	0.56	0.08	0.51	0.02	0.55	0.90	0	0.5	0.52	0.04	0.03	0.48	0.37	0.08	0.11	0.14	0.32	0.37	0.81	0.35	0.14

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

11	1.02	0.17	0.56	0.06	0.58	1.01	0.48	0.05	1.4	0.5	0	0.02	0.54	0.53	0.02	0.87	0.42	0.39	0.36	0.18	0.13	0.31	0.15	0.64
12	1.04	0.19	0.58	0.04	0.6	1.03	0.5	0.03	1.42	0.52	0.02	0	0.56	0.55	0.04	0.89	0.44	0.41	0.38	0.2	0.15	0.29	0.17	0.66
13	0.48	0.37	0.02	0.6	0.04	0.47	0.06	0.59	0.86	0.04	0.54	0.56	0	0.01	0.52	0.33	0.12	0.15	0.18	0.36	0.41	0.85	0.39	0.10
14	0.49	0.36	0.03	0.59	0.05	0.48	0.05	0.58	0.87	0.03	0.53	0.55	0.01	0	0.51	0.34	0.11	0.14	0.17	0.35	0.4	0.84	0.38	0.11
15	1	0.15	0.54	0.08	0.56	0.99	0.46	0.07	1.38	0.48	0.02	0.04	0.52	0.51	0	0.85	0.4	0.37	0.34	0.16	0.11	0.33	0.13	0.62
16	0.15	0.7	0.31	0.93	0.29	0.14	0.39	0.92	0.53	0.37	0.87	0.89	0.33	0.34	0.85	0	0.45	0.48	0.51	0.69	0.74	1.18	0.72	0.23
17	0.6	0.25	0.14	0.48	0.16	0.59	0.06	0.47	0.98	0.08	0.42	0.44	0.12	0.11	0.4	0.45	0	0.03	0.06	0.24	0.29	0.73	0.27	0.22
18	0.63	0.22	0.17	0.45	0.19	0.62	0.09	0.44	1.01	0.11	0.39	0.41	0.15	0.14	0.37	0.48	0.03	0	0.03	0.21	0.26	0.7	0.24	0.25
19	0.66	0.19	0.2	0.42	0.22	0.65	0.12	0.41	1.04	0.14	0.36	0.38	0.18	0.17	0.34	0.51	0.06	0.03	0	0.18	0.23	0.67	0.21	0.28
20	0.84	0.01	0.38	0.24	0.4	0.83	0.3	0.23	1.22	0.32	0.18	0.2	0.36	0.35	0.16	0.69	0.24	0.21	0.18	0	0.05	0.49	0.03	0.46
21	0.89	0.04	0.43	0.19	0.45	0.88	0.35	0.18	1.27	0.37	0.13	0.15	0.41	0.4	0.11	0.74	0.29	0.26	0.23	0.05	0	0.44	0.02	0.51
22	1.33	0.48	0.87	0.25	0.89	1.32	0.79	0.26	1.71	0.81	0.31	0.29	0.85	0.84	0.33	1.18	0.73	0.7	0.67	0.49	0.44	0	0.46	0.95
23	0.87	0.02	0.41	0.21	0.43	0.86	0.33	0.2	1.25	0.35	0.15	0.17	0.39	0.38	0.13	0.72	0.27	0.24	0.21	0.03	0.02	0.46	0	0.49
24	0.38	0.47	0.08	0.7	0.06	0.37	0.16	0.69	0.76	0.14	0.64	0.66	0.10	0.11	0.62	0.23	0.22	0.25	0.28	0.46	0.51	0.95	0.49	0

Table 2 is a Manhattan distance matrix that illustrates the level of difference or relative proximity between each pair of regencies/cities in South Sulawesi based on the Electronic-Based Government System (EBGS) achievement index. Each value in the matrix shows the total absolute difference between regions across all variables analyzed. The smaller the distance value between two regions, the more similar their characteristics or EBGS achievement levels are. Conversely, a larger distance value indicates that the two regions have significant differences in the implementation or achievement of EBGS indicators. After calculating the distance matrix, a clustering analysis will be performed using the Ward method with k selected as $k = 3$. The results of the analysis using the Ward method are presented in Table 3.

Table 3. Ward's Method Cluster Results

Number	Districts/Cities	Index EBGS	Cluster
1	Bantaeng	3.64	1
2	Barru	2.79	3
3	Bone	3.18	2
4	Bulukumba	2.56	3
5	Enrekang	3.20	2
6	Gowa	3.63	1

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

7	Jeneponto	3.10	2
8	Kepulauan Selayar	2.57	3
9	Makassar	4.02	1
10	Palopo	3.12	2
11	Parepare	2.62	3
12	Luwu	2.60	3
13	Luwu Timur	3.16	2
14	Maros	3.15	2
15	Pangkajene dan Kepulauan	2.64	3
16	Pinrang	3.49	1
17	Sindereng Rappang	3.04	2
18	Sinjai	3.01	2
19	Soppeng	2.98	2
20	Takalar	2.80	3
21	Tana Toraja	2.75	3
22	Toraja Utara	2.31	3
23	Wajo	2.77	3
24	Luwu Utara	3.26	2

The results of clustering using Ward's method produced three main clusters that describe the level of similarity between districts/cities in South Sulawesi Province based on the Electronic-Based Government System (EBGS) Index value. Ward's method works on the principle of minimizing the sum of squares within clusters, so that each cluster formed has a high level of internal homogeneity and clear heterogeneity between clusters. The results of this clustering can also be seen in the following dendrogram:

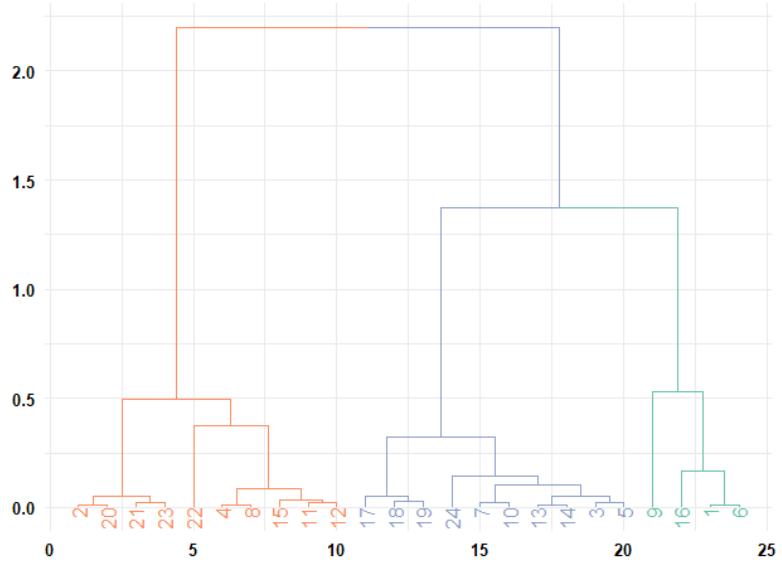


Fig 3. Dendrogram with Ward's Clustering Method

The results in Table 3 and Fig 3 identify the following characteristics: Cluster 1 (Green) consists of Bantaeng, Gowa, Makassar, and Pinrang Regencies, with relatively high

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

EBGS index values (ranging from 3.49 to 4.02) which are classified as Very Good. The regions in this group represent areas with the most optimal EBGS implementation in South Sulawesi, indicating mature, integrated, and efficient digital governance in the implementation of electronic-based public services. Cluster 2 (Blue) consists of regencies/cities such as Bone, Enrekang, Jeneponto, Palopo, East Luwu, Maros, Sidenreng Rappang, Sinjai, Soppeng, and North Luwu, with EBGS index scores ranging from 3.01 to 3.26, which is classified as Good. This group shows that the process of government digitization has been quite strong, although strengthening is still needed in terms of system integration, improving data interoperability, and optimizing the use of information technology in public services. Cluster 3 (Orange) includes districts/cities such as Barru, Bulukumba, Selayar Islands, Parepare, Luwu, Pangkajene and Islands, Takalar, Tana Toraja, North Toraja, and Wajo, with EBGS index scores ranging from 2.31 to 2.80, classified as Fair. This cluster represents regions with relatively limited adoption of EBGS, indicating the need for improvements in digital infrastructure, human resource capacity, and system integration.

Overall, this grouping pattern shows a gradation in the level of digital maturity of government in the South Sulawesi region. Regions with more complex government and public service activities, such as Makassar and Gowa, tend to have higher EBGS indices, while regions with limited geographical characteristics and resources show lower values. This also shows that the level of achievement in implementing the Electronic-Based Government System (EBGS) in the South Sulawesi region varies considerably between regencies/cities. The formation of three main clusters reflects differences in the level of digital maturity between regions, in terms of infrastructure, human resources, and electronic service management policies. These results reinforce that Ward's method is effective in identifying the uniformity of EBGS achievement levels between regions and can be the basis for formulating more targeted, evidence-based EBGS performance improvement policies and strategies that are in line with the specific conditions and needs of each region.

As a follow-up step, to ensure consistency of results and obtain a more comprehensive comparison, cluster analysis was also performed using a non-hierarchical method, namely K-Means Clustering. This method was chosen because it has the ability to optimize group formation based on the centroid by minimizing the distance between data in each cluster. Unlike Ward's hierarchical and agglomerative method, K-Means works through an iterative process of updating cluster centers, so that the grouping results can be more flexible to data variations and provide additional perspectives in seeing the patterns of EBGS index achievement between regions.

b. K-Means

The K-Means algorithm is used as an iterative clustering algorithm. This algorithm uses distance as a measurement standard, forms k clusters in a data set, calculates the average distance value, and then determines the initial centroid [10]. In this study, to group the districts/cities in South Sulawesi based on the Electronic-Based Government System (EBGS) service achievement index using the K-Means method, the initial step is to determine the number of clusters (k) to be formed. The optimal k value will be determined using three methods, namely the Elbow method, Silhouette method, and Gap Statistics method. The following are the results for each method:

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

- **Elbow Method**

When the number of clusters k is not predetermined by the application, we must choose that value; and this can be quite complicated. An elbow plot is a graph that displays the approximation error (SSE) on the y-axis against various values of k on the x-axis. The motivation behind the elbow criterion is based on the concept of diminishing returns, namely that as the value of k increases, the value of SSE will decrease after a certain point (Stop using the elbow criterion for k-means and how to choose the number of clusters instead) [17]. The following presents the results of the Elbow Plot to determine the optimal k in the clustering process:

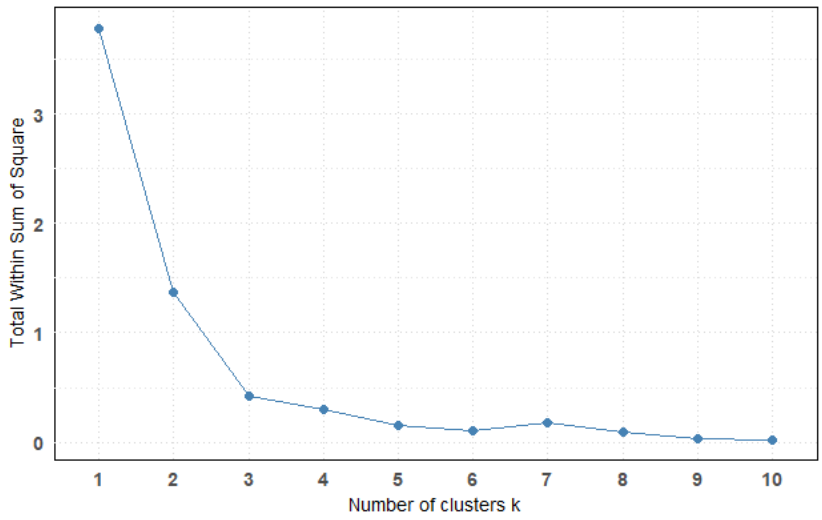


Fig 4. Elbow Plot

Fig 4 shows the Elbow Plot used to determine the optimal number of clusters (k) in the K-Means algorithm. The x-axis shows the number of clusters (k) tried, ranging from 1 to 10. The y-axis shows the Total Within Sum of Squares (SSE) value, which is the total variation in each cluster measured from the distance of the data points to their centroid. Based on Fig 4, it can be seen that the SSE value decreases sharply from $k = 1$ to $k = 3$, then the decrease slows down after $k = 3$. The SSE values are shown in Table 4 below:

Table 4. Total Within Sum of Squares (SSE)

k	1	2	3	4	5	6	7	8	9	10
SSE	3.78	1.37	0.42	0.30	0.16	0.11	0.18	0.10	0.03	0.02

Table 4 provides information that after $k = 3$, the decline in SSE values begins to slow down from 0.42, 0.30, 0.16, and so on. The sharp decline in SSE values at the beginning indicates that the addition of clusters at that stage significantly improves the quality of data separation. However, after a certain point (around $k = 3$), the decrease in SSE becomes insignificant, as seen in Fig 4, where the graph forms an "elbow" at $k = 3$. Therefore, it can be concluded that using the Elbow method, the optimal number of clusters (k) selected is $k = 3$. Next, we will discuss the determination of the optimal number of clusters using the Silhouette method.

- **Silhouette Method**

Most performance evaluation methods require training data (training set), but the Silhouette Index does not need it to assess clustering results. This makes the Silhouette Index

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

more suitable for use in clustering tasks or analyses [18]. In a study conducted by [19], a Silhouette Coefficient-based Weighted K-Means algorithm was proposed that automatically adjusts feature weights during the clustering process. In this algorithm, the process of searching for optimal clusters is converted into an optimization problem that aims to maximize the Silhouette Coefficient value of the resulting clusters. The weight of each feature is updated iteratively during the K-Means process. The following presents the Silhouette Coefficient Plot results to determine the optimal k in the clustering process:

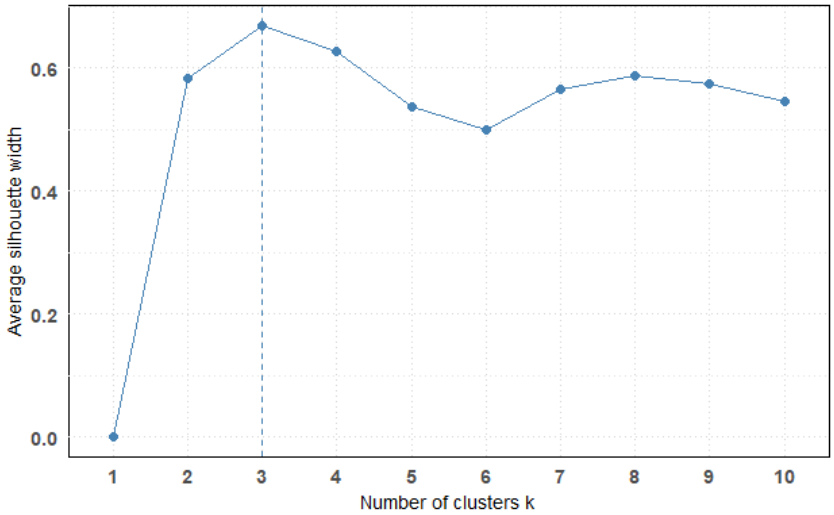


Fig 5. Silhouette Coefficient Plot

Fig 5 shows the Silhouette Coefficient Plot, which is used to determine the optimal number of clusters (k) in the K-Means algorithm based on the average Silhouette coefficient value. The x-axis shows the number of clusters (k) tested, from 1 to 10. The y-axis shows the average Silhouette Coefficient value, which measures how well each object is placed in its cluster. The value ranges from -1 to 1, where a value close to 1 indicates that the data fits very well with its own cluster and is far from other clusters, a value close to 0 indicates that the data is on the border between two clusters, and a negative value indicates the possibility that the data is placed in the wrong cluster. The average Silhouette Coefficient values are presented in Table 5 below:

Table 5. Average Silhouette Width

k	1	2	3	4	5	6	7	8	9	10
Average Silhouette	0.00	0.58	**0.67**	0.63	0.54	0.50	0.57	0.59	0.57	0.54

Fig 5 and Table 5 provide information that the highest Silhouette value occurs at $k = 3$, with an average of around 0.67. This indicates that the formation of three clusters ($k = 3$) provides the best and most stable cluster separation compared to other cluster numbers. After $k > 3$, the Silhouette value begins to decline and fluctuate, indicating that adding more clusters does not improve the separation quality, but rather tends to worsen it. Therefore, it can be concluded that with the Silhouette method, the optimal number of clusters (k) selected is $k = 3$, because at that point the highest average Silhouette value is obtained, which indicates the best and most representative cluster structure for the data pattern. Next, the determination of the optimal number of clusters using the Gap Statistic method will be discussed again.

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

- **Gap Statistics Method**

The Gap Statistic method is used to determine the optimal number of clusters by comparing the level of variation within clusters at various cluster numbers (k) with the expected value if the data had no cluster structure (based on the reference distribution). This approach helps identify the point at which increasing the number of clusters no longer provides a significant improvement in data separation [20]. The following presents the results of the Gap Statistic Plot to determine the optimal k in the clustering process:

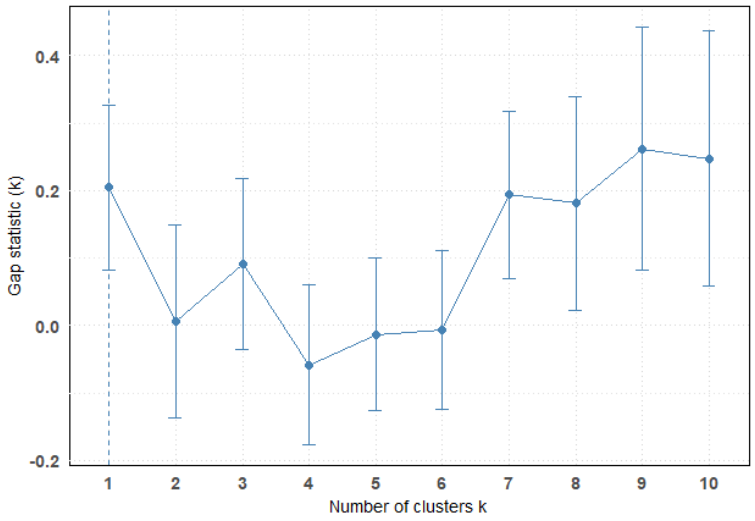


Fig 6. Gap Statistic Plot

Fig 6 shows the Gap Statistic Plot, which is used to determine the optimal number of clusters (k) in the K-Means algorithm. The x-axis shows the number of clusters (k) tested, ranging from 1 to 10. The y-axis shows the Gap Statistic value ($G(k)$), which is obtained from the difference between the logarithm of the variation within clusters (within-cluster variation) in the original data and its expected value from the reference data (null reference distribution). The larger the Gap Statistic value, the better the cluster structure formed, because it indicates that the cluster formation is clearer than random data without structure. The Gap Statistic values are presented in Table 6 below:

Table 6. Gap Statistics

k	1	2	3	4	5	6	7	8	9	10
log (W_k)	0.97	0.35	-0.21	-0.41	-0.77	-1.05	-1.47	-1.68	-1.98	-2.19
E [log (W_k)]	1.18	0.36	-0.12	-0.47	-0.78	-1.05	-1.28	-1.50	-1.71	-1.94
Gap Statistic ($G(k)$)	0.20	0.01	0.09	-0.06	-0.01	-0.01	0.19	0.18	0.26	0.25

Note: log (W_k) is the logarithm of Within-Cluster Dispersion, E [log (W_k)] is the expectation of log (W_k).

Fig 6 and Table 6 provide information that the highest gap value is at k = 1 (0.20). After that, the gap value fluctuates around zero and some are even negative. This indicates that the cluster structure in the data is not very strong, the data tends to be homogeneous, and the Gap Statistic method identifies k = 1 as the optimal number of clusters.

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

Based on the results of determining the optimal number of clusters using three methods, varying results were obtained. The Elbow and Silhouette Coefficient methods consistently showed that the optimal number of clusters was $k = 3$, which was indicated by a clear elbow point on the Elbow graph and the highest Silhouette coefficient value at $k = 3$. Meanwhile, the Gap Statistic method showed optimal results at $k = 1$, which theoretically indicated that the cluster structure in the data was relatively weak or that the data was homogeneous. However, considering the purpose of the analysis and the clearer pattern of variation between observations in the Elbow and Silhouette results, this study determined that the optimal number of clusters to be selected is $k = 3$. This selection is considered the most representative in describing the differences in characteristics between data groups. Next, the clustering process was carried out with K-Means using $k = 3$. The clustering results are presented in Fig 7 below:

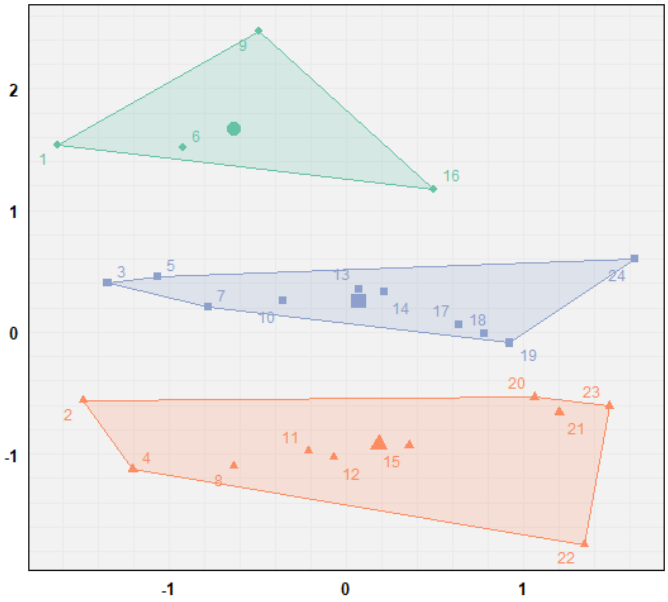


Fig 7. K-Means Clustering Plot

Based on the results of the K-Means Clustering analysis shown in Fig 7, the districts/cities in South Sulawesi Province were divided into three clusters with 4, 10, and 10 regions, respectively. These results indicate that most of the regencies/cities in South Sulawesi are in the Good category: Cluster 2 (Blue) and Fair category: Cluster 3 (Orange), while only a small portion of the regions are classified as Very Good: Cluster 1 (Green). The complete results are presented in Table 7:

Table 7. K-Means Clustering

Cluster	Number of Districts/Cities	Average EBGS	Within-Cluster Sum of Squares (WSS)	Category
1	4	3.70	0.15	Very Good
2	10	3.12	0.07	Good
3	10	2.64	0.20	Fair

Table 7 provides information that the average EBGS index value shows a clear difference between clusters. Cluster 1 has the highest average of 3.70, so it is categorized as an area

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

with a very good level of EBGS service achievement. Cluster 2, with an average of 3.12, is categorized as good, while Cluster 3, with an average of 2.64, is categorized as fair. The Within-Cluster Sum of Squares (WSS) value also provides an overview of the level of homogeneity between regions within each cluster. Cluster 2 has the smallest WSS value (0.07), indicating that the characteristics of the regions in this group are relatively uniform. In contrast, Cluster 3 has the highest WSS value (0.20), indicating greater variation between regions in that group. Meanwhile, overall, the proportion of Between-Cluster Variation to total variation is 88.8%, indicating that this clustering model is able to separate regions well into different groups.

Overall, the K-Means clustering results indicate that three clusters are sufficient to describe the EBGS achievement patterns in South Sulawesi. Cluster 1 represents regions with very high EBGS performance, Cluster 2 indicates moderate performance, and Cluster 3 reflects relatively lower EBGS maturity levels. The clustering structure confirms a clear stratification of digital governance maturity across districts/cities.

From a policy perspective, regions in Cluster 3 should be prioritised for targeted interventions, including strengthening digital infrastructure, enhancing human resource capacity, and improving system interoperability. Cluster 2 regions may benefit from consolidation and optimisation strategies to improve service integration and digital performance, while Cluster 1 regions can serve as benchmark models for best practices in EBGS implementation and digital governance maturity.

4.4 Cluster Evaluation

At this stage, the clustering results obtained from the K-Means and Ward's Linkage methods are evaluated by assessing the validity and stability of the clusters formed. The evaluation is carried out to determine the extent to which the two methods are able to produce statistically valid groupings that are stable against method variations. The validity of the clusters is assessed using three main measures, namely the Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index, which complementarily assess the quality of separation and compactness between clusters. Meanwhile, cluster stability is analyzed by comparing the suitability of the grouping results between the two methods using the Adjusted Rand Index (ARI). These three validity measures and one stability measure are used to provide a comprehensive overview of the quality and consistency of the clustering results. A summary of the comparison of the index values for the K-Means and Ward's Linkage methods is presented in Table 8 below:

Table 8. Validity and Stability of Cluster Results

Validity Measure			Stability Measure
Silhouette Coefficient	Davies-Bouldin Index (DBI)	Calinski-Harabasz Index (CHI)	Adjusted Rand Index (ARI)
0.67	0.39	83.02	1.00

For transparency and reproducibility, the internal validity indices are reported for both Ward's linkage and K-Means clustering, although both methods yielded identical numerical values, as shown in Table 8. The internal validity indices were computed separately for Ward's linkage and K-Means clustering, and both methods yielded identical values for the Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index, indicating consistent within-cluster compactness and between-cluster separation across

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

clustering approaches. This evaluation aims to ensure that the clustering results truly represent natural patterns in the EBGS Index data between districts/cities in South Sulawesi, and have a high level of consistency with variations in the analysis method. Based on the calculations in Table 8, the K-Means method produced a Silhouette value of 0.67, a DBI value of 0.39, and a CHI value of 83.02. The relatively high Silhouette value indicates that the objects in the cluster have strong internal similarities and are clearly separated from other clusters. The low DBI value reinforces the evidence that the clusters have minimal overlap, while the high CHI value indicates a good balance between internal compactness and separation between clusters. In this study, a high CHI value indicates that differences in EBGS maturity levels between clusters are substantial, while districts/cities within the same cluster exhibit relatively homogeneous EBGS performance. Interestingly, the Ward's Linkage method produced validity index values identical to the K-Means method for all three measures. This similarity not only shows that both methods form the same cluster structure, but also confirms that the data grouping pattern is very stable and robust against differences in algorithmic approaches. In other words, both hierarchical (Ward's Linkage) and non-hierarchical (K-Means) approaches reveal a consistent and statistically robust natural structure of the data. The stability of the clusters was then evaluated using the Adjusted Rand Index (ARI) to measure the similarity between the results of the two methods. The ARI value of 1.00 indicates a very high level of similarity between the clustering results of the two methods, confirming the robustness and stability of the identified cluster structure. Overall, the similarity of all validity index values and the perfect ARI value indicate that the clustering results obtained are highly reliable, valid, and stable, with a robust grouping structure against variations in the analysis method. This confirms that the grouping of districts/cities based on the EBGS Index in South Sulawesi has succeeded in revealing clear, consistent, and representative patterns of the actual conditions in the field.

The identical clustering solutions produced by Ward's linkage and K-Means indicate that the EBGS index exhibits a well-defined and low-ambiguity cluster structure in a univariate feature space. Given that the analysis is based on a single composite EBGS index, the partitioning problem becomes relatively constrained, leading both hierarchical and partition-based algorithms to converge to the same optimal grouping. Furthermore, hierarchical clustering results are commonly used as a reference framework for centroid initialization or validation in partitioning algorithms, which may contribute to the convergence of K-Means to the same cluster configuration identified by Ward's method.

5. CONCLUSION

This study shows that Ward's Linkage and K-Means methods produce identical cluster patterns for the EBGS Index of districts/cities in South Sulawesi, indicating that the EBGS data structure has high statistical coherence and consistent stability of results despite differences in analytical approaches. Statistically, this reflects that the grouping patterns between regions are strong and insensitive to method variations, thus providing confidence in the reliability of the grouping model formed.

These findings also indicate that the level of development of electronic-based government systems in the region shows a statistically uniform distribution pattern, with a tendency toward homogeneity in the level of digital maturity between districts/cities. For local governments, these results can be used as a basis for formulating policies and strategies to strengthen data-based digital governance, especially in directing EBGS development and

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

performance improvement programs in regions included in clusters with relatively low achievements.

However, this study has several limitations, including the use of a limited number of variables and not considering external factors that may affect EBGS achievements, such as human resources, digital infrastructure, or regional socioeconomic conditions.

Therefore, further research is recommended to use more statistically complex clustering approaches such as fuzzy clustering or model-based clustering, as well as adding supporting variables to obtain a more comprehensive understanding of the dynamics of EBGS development at the regional level.

REFERENCES

[1] A. Jaeger and D. Banks, "Cluster analysis: A modern statistical review," *WIREs Computational Statistics,* vol. 15, no. 3, pp. 1-17, 2022.

[2] A. A. Ismael, . Z. A. A. AL-Bairmani and I. F. F. AL-Masferi , "Comparison of the Two Methods of Cluster Analysis (Non-Hierarchical and Hierarchical) in the Classification of Laboratory Quality (GLP) at the University of Babylon," *Journal of University of Babylon for Pure and Applied Sciences,* vol. 29, no. 3, pp. 34-43, 2021.

[3] I. O. Agada, O. Peter and E. J. Eweh, "Hierarchical and Non-Hierarchical Cluster Classification of Precipitation Time Series Data in Nigeria," *NIGERIAN JOURNAL OF THEORETICAL AND ENVIRONMENTAL PHYSICS ,* vol. 2, no. 1, pp. 36-48, 2024.

[4] A. L. Adu, R. Hartanto and . S. Fauziati, "HAMBATAN-HAMBATAN DALAM IMPLEMETASI LAYANAN SISTEM PEMERINTAHAN BERBASIS ELEKTRONIK (EBGS) PADA PEMERINTAH DAERAH," *Jurnal Informatika dan Komputer,* vol. 5, no. 3, pp. 215-223, 2022.

[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," *Eigen Mathematics Journal ,* vol. 4, no. 1, pp. 8-17, 2021.

[6] N. Randriamihamison, N. Vialaneix and P. Neuvial, "Applicability and Interpretability of Hierarchical Agglomerative Clustering With or Without Contiguity Constraints," *Journal of Classification,* vol. 38, no. 2, pp. 363-389, 2021.

[7] M. Wu, S. Zhang, R. Sun, J. Tang, S. Hu and F. Zhang, "Anomaly Detection Method for Lithium-Ion Battery Cells Based on Time Series Decomposition and Improved Manhattan Distance Algorithm," *American Chemical Society Omega,* vol. 9, no. 2, pp. 2409-2421, 2024.

[8] M. . F. Nur and A. Siregar, "Exploring the Use of Cluster Analysis in Market Segmentation for Targeted Advertising," *IAIC Transactions on Sustainable Digital Innovation ,* vol. 5, no. 2, p. 158–168, 2024.

[9] M. Vichi, C. Cavicchia and . P. J. F. Groenen, "Hierarchical Means Clustering," *Journal of Classification,* vol. 39, p. 553–577, 2022.

[10] B. Chong , "K-means clustering algorithm: a brief review," *Academic Journal of Computing & Information Science,* vol. 4, no. 5, pp. 37-40, 2021.

[11] E. L. Ofetotse, E. A. Essah and R. Yao, "Evaluating the Determinants of Household Electricity Consumption Using Cluster Analysis," *Journal of Building Engineering,* vol. 43, 2021.

[DOI: 10.26714/jsunimus.14.1.2026.01-21]

[12] L. Lenssen and E. Schubert, "Medoid Silhouette clustering with automatic cluster number selection," *Information Systems,* vol. 120, 2024.

[13] A. J. Jacob, I. J. Daniel and C. S. Aneke, "Determination Of Optimal Number Of Clusters Using Gap Statistics And Elbow Methods," *International Multilingual Journal of Science and Technology,* vol. 9, no. 3, pp. 7361-7366, 2024.

[14] A. Goldstein, Y. Shahar, M. W. Raymond, H. Peleg, E. Ben-Chetrit, A. Ben-Yehuda, . E. Shalom, . C. Goldstein, S. S. Shiloh and G. Almoznino, "Multi-Dimensional Validation of the Integration of Syntactic and Semantic Distance Measures for Clustering Fibromyalgia Patients in the Rheumatic Monitor Big Data Study," *bioengineering,* vol. 11, no. 1, pp. 1-31, 2024.

[15] F. Murtagh and P. Legendre, "Ward's Hierarchical Agglomerative Clustering Method:Which Algorithms Implement Ward's Criterion?," *Journal of Classification,* pp. 274-295, 2014.

[16] V. Kathiresan and N. A. Vasanthi, "OUTLIER DETECTION ON FINANCIAL CARD OR ONLINE TRANSACTION DATA USING MANHATTAN DISTANCE BASED ALGORITHM," *International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ,* vol. 2, no. 12, pp. 1100-1103, 2016.

[17] E. Schubert, "Stop using the elbow criterion for k-means and how to choose the number of clusters instead," *ACM SIGKDD Explorations Newsletter,* vol. 25, no. 1, pp. 36-42, 2023.

[18] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," *Entropy,* vol. 23, no. 6, pp. 1-17, 2021.

[19] H. Lai, T. Huang, B. Lu, S. Zhang and R. Xiaog, "Silhouette coefficient-based weighting k-means algorithm," *Neural Computing and Applications,* vol. 37, no. 5, pp. 3061-3075, 2025.

[20] A. Abdulhafedh, "Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation," *Journal of City and Development,* vol. 3, no. 1, pp. 12-30, 2021.

[DOI: 10.26714/jsunimus.14.1.2026.01-21]