Jurnal Pendidikan Matematika (Kudus)
P-ISSN 2615-3939 | E-ISSN 2723-1186
https://journal.iainkudus.ac.id/index.php/jmtk
DOI: http://dx.doi.org/10.21043/jpmk.v9i1.34232
Volume 9, Number 1, June 2026, pp. 27-42

Comparison between Average Linkage and K-Means Clustering Methods in Grouping Regencies or Cities by Open Unemployment Rate in Central Java

Albertus Dion Sarah[1], Supriadin[1], Karin Dita Widiyanti[1], Jesicha Roselina Arsusma[1], Yolan Triky [1], Prizka Rismawati Arum[1*]
[1] Departement of Statistics, Universitas Muhammadiyah Semarang, Semarang, Indonesia
**Correspondence: prizka.rismawatiarum@unimus.ac.id*

Abstract

Unemployment is a serious social and economic problem that hinders growth, including in Central Java Province. Based on the National Labor Force Survey (Sakernas), the unemployment rate in Central Java Province reached 5.96% in February 2021, an increase of 1.76% compared to February 2020. This study aims to compare two clustering methods, namely Average Linkage and K-Means Clustering, in grouping regions based on three indicators: population size, regencies/cities Minimum Wage (UMK), and labor force size. Evaluation was performed using the Silhouette Coefficient value to assess cluster quality. The analysis results indicated that the K-Means Clustering method with 2 clusters is the most optimal approach for classifying the open unemployment rate in Central Java Province. This is demonstrated by a higher Silhouette Coefficient value of 0.676 for K-Means Clustering, compared to 0.444 for the Average Linkage method. This clustering result divides the region into 2 clusters, namely the first cluster consisting of 8 regencies/cities and the second cluster comprising 27 regencies/cities. These findings are expected to assist local governments in formulating targeted unemployment mitigation policies.

Keywords: Average Linkage, Clustering, K-Means, Silhouette coefficient, Unemployment

Abstrak

Pengangguran merupakan masalah sosial dan ekonomi serius yang menghambat pertumbuhan, termasuk di Provinsi Jawa Tengah. Berdasarkan Survei Angkatan Kerja Nasional (Sakernas), mencatat tingkat pengangguran di Provinsi Jawa Tengah mencapai 5.96% pada Februari 2021, atau meningkat 1,76% dibandingkan Februari 2020. Penelitian ini bertujuan membandingkan dua metode pengelompokan, yaitu *Average Linkage* dan *K-Means Clustering*, dalam mengelompokkan wilayah berdasarkan tiga indikator: jumlah penduduk, Upah Minimum Kabupaten/Kota (UMK), dan jumlah angkatan kerja. Evaluasi dilakukan menggunakan nilai *Silhouette*

Coefficient untuk menilai kualitas *cluster*. Hasil analisis menunjukkan bahwa metode *K-Means Clustering* dengan jumlah 2 *cluster* merupakan pendekatan yang paling optimal dalam mengelompokkan tingkat pengangguran terbuka di Provinsi Jawa Tengah. Hal ini ditunjukkan oleh nilai *Silhouette Coefficient* yang lebih tinggi, yaitu 0.676 untuk *K-Means Clustering*, dibandingkan dengan 0.444 pada metode *Average Linkage*. Hasil klasterisasi ini membagi wilayah menjadi 2 *cluster*, yaitu *cluster* pertama yang terdiri dari 8 Kabupaten/Kota dan *cluster* kedua yang mencakup 27 Kabupaten/Kota. Temuan ini diharapkan dapat membantu pemerintah daerah dalam menyusun kebijakan penanggulangan pengangguran secara lebih tepat sasaran.

Kata Kunci: *Average Linkage, K-Means, Clustering, Silhouette Coefficient*, Pengangguran

Introduction

Unemployment is a major socio-economic issue that affects economic growth and social welfare, including in Central Java Province. It refers to individuals who are not working but are actively seeking employment (Hegelund & Taalbi, 2023). High unemployment rates indicate that human resources are not utilized optimally, which can lead to reduced income levels and increased socio-economic inequality (Azhar et al., 2024). In Indonesia, unemployment is influenced by various factors, including population growth, imbalance between labor supply and demand, low workforce adaptability to technological developments, education levels, and regional disparities in minimum wages (Kurniawan, 2024). Limited job opportunities for job seekers can cause various problems in a region, such as a decline in welfare, work productivity, and community income. Further impacts include an increase in poverty, crime rates, and disparities in living standards (Doni et al., 2023).

Unemployment is a common problem faced by many in Indonesia. Based on the latest data from the Central Statistics Agency (BPS) in 2022, the open unemployment rate in Indonesia was recorded at 5.86% of the total population (Maisuun & Darmawanti, 2023). Although this shows a decline from 2020 to 2021, the unemployment rate in Indonesia is still high. In August 2020, the unemployment rate reached 7.07 percent. A year later, in August 2021, the figure fell by 0.58% to 6.49%. Based on the National Labor Force Survey (Sakernas) in Central Java Province, the unemployment rate in February 2021 was recorded at 5.96%, an increase of 1.76% compared to February 2020 (Laber & Batista, 2024). High unemployment rates are a significant social and economic problem, as the increase in the number of unemployed people in Indonesia hinders economic growth, including in Central Java (Frisnoiry et al., 2024).

Clustering was conducted to identify hidden patterns and characteristics in open unemployment data in various regencies/cities in Central Java Province. With clustering, areas with similar unemployment conditions can be grouped together, facilitating more in-depth analysis and supporting the formulation of more targeted policies (Zubair et al., 2022). A comparison of clustering methods using hierarchical and non-hierarchical approaches was conducted through an analysis of two main methods, namely Average Linkage and K-Means Clustering. Average Linkage is an algorithm in hierarchical clustering that groups data based on the average distance between objects in each cluster (Butar Butar, 2023). Meanwhile, K-Means is a clustering method that divides data into a predetermined number of clusters, based on the degree of similarity between objects in each cluster (Abdulhafedh, 2025) (Messakh et al., 2023)

The performance of these clustering methods is evaluated using the silhouette coefficient, which measures the similarity of an object to its own cluster compared to other clusters. The silhouette value ranges from -1 to 1, where higher values indicate better clustering quality (Paembonan & Abduh, 2021). The case study used for this analysis is the open unemployment rate in Central Java.

Several previous studies have applied clustering methods to unemployment-related problems. For instance (Arifatul Ulya et al., 2023) sed the K-Means method to cluster unemployment rates in Central Java, and produced four clusters with varying regional distributions. However, the study only applied a single method without evaluating whether the selected method was optimal. Furthermore, research conducted by (Hasmira et al., 2023) compared several hierarchical clustering methods and found that Average Linkage performed best based on cophenetic correlation. Nevertheless, the study focused on poverty indicators in South Sulawesi, making its findings less directly applicable to unemployment cases in Central Java.

These studies indicate that while clustering has been widely applied, most research either focuses on a single method or examines different case studies, resulting in limited insight into the comparative performance of clustering methods for unemployment analysis in Central Java (Nugroho et al., 2024). In addition, previous studies rarely evaluate clustering quality using robust validation measures such as the silhouette coefficient.

Based on this gap, this study aims to compare the performance of Average Linkage and K-Means clustering methods in grouping regencies/cities based on

unemployment-related indicators in Central Java Province, and to determine the most optimal method for producing accurate and meaningful clusters

Method

Data Source

The data used in this research is secondary data collected from other available sources. The data was obtained from the Central Statistics Agency (Badan Pusat Statistik, BPS) of Central Java and can be accessed through the website www.bps.go.id. The research variables used are population size, Regional/City Minimum Wage (UMK) amount, and labor force size that influence the open unemployment rate based on regions/cities in Central Java Province. These three variables are further defined in Table 1.

Table 1. Variable Definitions

Variable	Definition
Population Total	The total population living in the Regency/City in Central Java.
UMK Amount	The minimum wage amount set by the Regency/City in Central Java.
Labor Force Total	The number of individuals who are ready and able to work in the labor force.

Research Procedure

The research procedures taken to cluster the open unemployment rate in Central Java based on regencies/cities are as follows:

1. Performing initial data exploration to ensure the completeness and reliability of the data.
2. Performing descriptive statistics for each independent variable.
3. Performing the data standardization process which carried out to standardize the range of values for all variables using the z-score method, which is expressed by the following equation (Putra et al., 2024).

$$Z = \left(\frac{x - \mu}{\sigma}\right) \quad (1)$$

4. Calculating the distance between each data from each cluster using the Euclidean formula as follows (Xu et al., 2021):

$$d_{(xi,\mu j)} = \sqrt{\sum_{k=1}^{p} (x_{ik} - \mu_{jk})^2} \quad (2)$$

$d_{(xi,\mu j)}$ = distance between data point $i$and centroid of cluster j

μ_j = *centroid of the j-th cluster*

x_{ik} = *value of the k-th variable for data point i*

μ_{jk} = *value of the k-th variable for centroid j*

P = *number of variables (dimensions)*

k = *index of variables, where* $k = 1,2,\ldots,p$

5. Performing the Clustering Process using the Average Linkage Method.
   1) One approach in the Clustering method is Average Linkage (Sinta Amelia, 2022). This method calculates the average distance between all pairs of objects in one cluster and all pairs of objects in a different cluster (Syalomitha R et al., 2021). This average calculation uses the following formula:

$$d_{(AB)k} = \frac{\Sigma_u \Sigma_v d_{uv}}{n_{(AB)} n_c} \quad (3)$$

   According to Johnson (2007):

   d_{uv} = is the distance between object u in cluster (AB) and object v in cluster C.

   $n_{(AB)}$ and n_C = are the number of objects in groups (AB) and (C).

   2) Determining the optimal number of clusters (k) using the dendrogram method (Xu et al., 2021).
   3) After obtaining the optimal number of clusters, the clusters are formed, and a cluster characteristic analysis of the open unemployment rate based on regencies/cities in Central Java Province is performed.

6. Performing the Clustering Process using the K-Means Clustering Method
   1) The K-Means method divides data into several clusters, where data with similar characteristics are grouped into one cluster, while data with different characteristics are placed into different clusters (Azzahra & Amru Yasir, 2024). The K-Means algorithm iteratively assigns each data point to the nearest centroid based on the Euclidean distance and updates the centroid until convergence is achieved. The objective function of K-Means is defined as:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} || x_i - c_j || \quad (4)$$

   Where: xi is the i-th data point (criteria data) and μj is the centroid of the j-th cluster.

2) Determining the optimal number of clusters (k) using the silhouette method (Yusniyanti et al., 2021)
3) After determining the optimal number of clusters, the clusters are formed, and a cluster characteristic analysis of the open unemployment rate based on regencies/cities in Central Java Province is performed.

7. Comparing the evaluation results of the clustering methods using Hierarchical Clustering and K-Means based on the silhouette coefficient value. If the average silhouette value approaches 1, it indicates better cluster quality. Conversely, if the average silhouette value approaches -1, the cluster quality is considered poor (Iqbal & Nurul Huda, 2023). The measurement criteria for the silhouette coefficient value can be seen in Table 2.

Table 2. Silhouette Coefficient Criteria

Silhouette Coefficient	Evaluation Criteria
0.71< SC < 1.0	The resulting structure is strong
0.51 < SC < 0.70	The resulting structure is good
0.26 < SC < 0.50	The resulting structure is weak
SC ≤ 0.25	Unstructured

Results

This study produced findings obtained through the application of descriptive analysis and cluster analysis methods.

Descriptive Analysis

The factors influencing the open unemployment rate in Central Java Province can be observed in Table 3.

Table 3. Results of Descriptive Analysis of Three Indicators

Total Population		UMK		Total Workforce	
Minimum	122105	Minimum	1958170	Mininum	69199
1st Qu.	841539	1st Qu.	2041991	1st Qu.	488949
Median	1047226	Median	2138248	Median	583820
Mean	1024737.6	Mean	2198563	Mean	601975
3rd Qu.	1345010	3rd Qu.	2277327	3rd Qu.	801679
Maximum	2043077	Maximum	3060349	Maximum	1046667

Table 3 presents the Regional/City Minimum Wage (UMK) data in Central Java Province for 2023, which had the highest average value of IDR 2.198.563, reflecting the average minimum wage applicable across all regencies/cities. The average population size

was recorded at 1.024.737,6 million people, indicating the population distribution in Central Java Province. The average labor force size was 601.975 people, which describes the number of available workers in the regencies/cities of Central Java Province.

Standardized Data

As a preparatory step before the clustering process, all unemployment rate variables are standardized using the z-score method to ensure uniform data scaling and to avoid the influence of differences in measurement units across variables. The results of the z-score standardization are presented in Table 4.

Table 4. Z-Score Standardized Data

No	Regency or City	X_1	X_2	X_3
1	Kabupaten Cilacap	-0.95	1.28	0,80
2	Kabupaten Banyumas	-0.12	1.86	-0,38
⋮	⋮	⋮	⋮	⋮
34	Kota Pekalongan	1.38	-1,45	0,42
35	Kota Tegal	2.25	-1,47	-0,23

Based on Table 4, the standardized z-score results for all poverty indicator variables in regencies/cities in Central Java Province ranged from -2.39 to 4.02. This range reflects significant differences in socio-economic characteristics between regions.

Euclidean Distance

After standardizing the data using the z-score method, the next step is to calculate the Euclidean distance. The distance of each regency/city to each cluster is computed to assess regional proximity. A smaller distance indicates a higher similarity in characteristics. The results of the distance calculations to Cluster 1 (C_1) and Cluster 2 (C_2) are presented in Table 5.

Tabel 5. Euclidean Distance

Objek	C_1	C_2	Objek	C_1	C_2	Objek	C_1	C_2
1	3,11	1,55	13	1,53	1,45	25	2,10	208
2	3,32	1,43	14	3,22	1,37	26	0,98	1,82
3	2,78	0,66	15	2,66	0,97	27	3,37	1,62
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	1,31	2,12	23	2,51	2,14	35	1,99	3,48
12	3,25	1,40	24	1,47	2,27			

Based on Table 5, the results of the Euclidean distance calculation, objects with a smaller distance value from the center of Cluster 1 (C_1) are grouped into Cluster 1, such as Object 11 (1.31 < 2.12). Conversely, objects that have a smaller

distance from the center of Cluster 2 (C_2) tend to be classified into Cluster 2, such as Object 3 (0.66 < 2.78) and Object 12 (1.40 < 3.25).

Clustering Using the Average Linkage Method

Before conducting cluster analysis, the first step is to determine the optimal number of clusters using the dendrogram method, as shown in Figure 1.

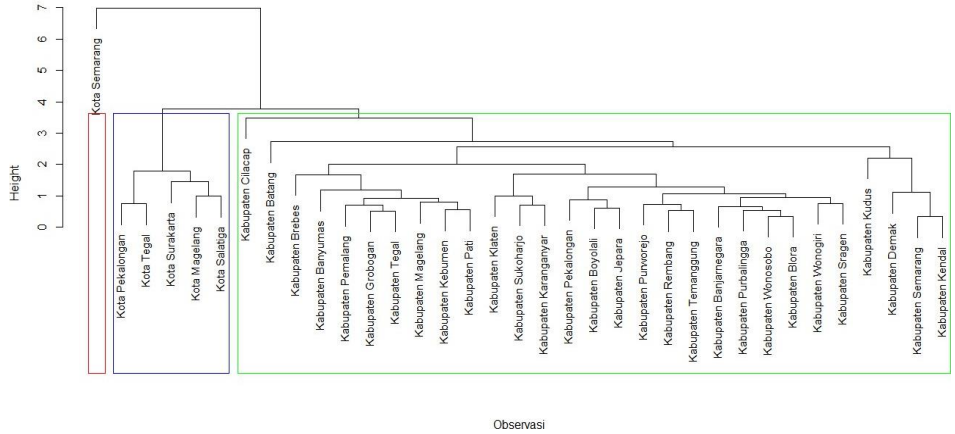

Figure 1. Average Linkage Cluster Dendrogram

Based on Figure 1, the initial number of clusters is determined by cutting branches at a certain distance level, resulting in clearly defined cluster groups. However, considering that determining the number of clusters based on a dendrogram tends to be subjective, additional, more objective considerations are needed. Therefore, the selection of three clusters is not only based on the dendrogram visualization, but also considering the level of distance difference between the cluster merging processes (linkage distance), which indicated a significant spike before the formation of the three clusters. Based on these results, three clusters were obtained with details: the first cluster consisted of 29 districts/cities, the second cluster consisted of 5 districts/cities, and the third cluster consisted of only 1 district/city. The composition of the third cluster can be seen in Table 6.

Table 6. Clustering Results Using the Average Linkage Method

Cluster	Regencies/Cities	Results Identification
1	Cilacap, Banjarnegara, Purbalingga, Pemalang, Tegal, Brebes, Batang, Pekalongan, Kendal, Temanggung, Semarang, Demak, Jepara, Kudus, Pati, Blora, Grobogan, Rembang, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Purworejo, Kebumen, Banyumas, and Wonosobo Regencies.	Consisting of regions with a high population and labor force, but with the Regency/City Minimum Wage (UMK) at a medium level. The large population and labor force, accompanied by limited employment absorption capacity, contribute to a higher open unemployment rate.
2	Pekalongan City, Tegal City, Magelang City, Salatiga City, and Surakarta City.	These cities tend to have a more advanced economic structure and better employment opportunities, resulting in a more controlled unemployment rate.
3	Semarang City.	Semarang City as the economic center of Central Java, has the highest population, labor force, and Regency/City Minimum Wage (UMK) among all clusters. With greater employment opportunities, the city also has a lower open unemployment rate.

After the clustering process is complete, the next step is to perform a characteristic analysis for each cluster to determine the average value for each variable. The characteristics of each cluster are presented in Table 7.

Table 7. Cluster Characteristics Based on Mean Values

Variable	Cluster 1	Cluster 2	Cluster 3
Total Population	1157583	120231	1694743
Total Minimum Weight (UMK)	2169453	2195038	3060349
Total Workforce	666886	160086	929014

After performing Clustering on the open unemployment rate in Central Java Province in 2023 based on regencies/cities, the characteristic analysis showed that Cluster 1, consisting of 29 regencies/cities, had an average population of 1.157.583 million people, an average regional/city Minimum Wage (UMK) of IDR 2.169.453,

and an average labor force size of 666.886 people. Cluster 2, comprising 5 cities, had an average population of 120,231 million people, an average UMK of IDR 2.195.038, and an average labor force size of 160.086 people. Meanwhile, Cluster 3, which consists only of Semarang City, had a population of 1.694.743 million people, an UMK of IDR 3.060.349, and a labor force size of 929.014 people. This data demonstrates significant variation in characteristics among the clusters, which can be used as a basis for policies and strategies in managing the open unemployment rate in Central Java Province.

K-Means Clustering

The first step in cluster analysis is to determine the optimal number of clusters using the silhouette method. In this analysis, the value of k was tested from k = 2 to k = 10 to observe changes in the average silhouette value, as shown in Figure 2.

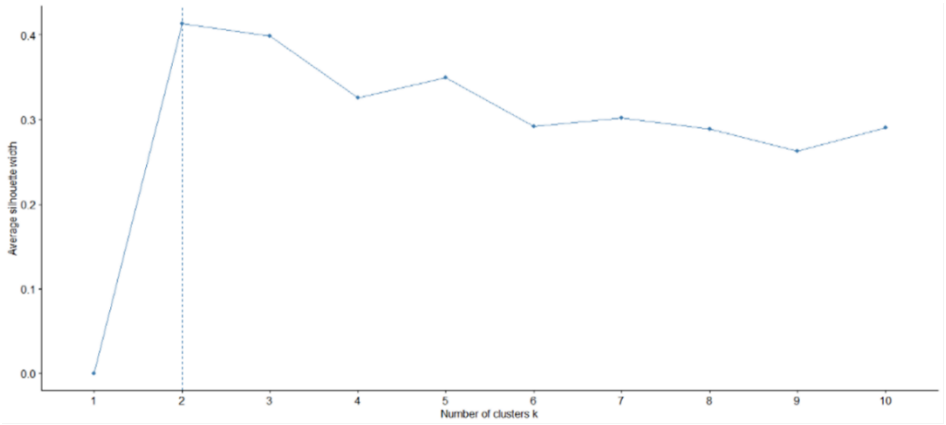

Figure 2. Visualization of the Optimal k Cluster Using the Silhouette Method

Based on Figure 2, the optimal k value for the K-Means method is k = 2, as the average silhouette value reaches its peak at this point. This indicates that the clusters at k = 2 have the clearest and most well-defined separation compared to other cluster numbers. Therefore, it can be concluded that the optimal number of clusters to be formed is two. The clustering results of regencies/cities based on the open unemployment rate in Central Java Province using the K-Means Clustering method can be seen in Figure 3.

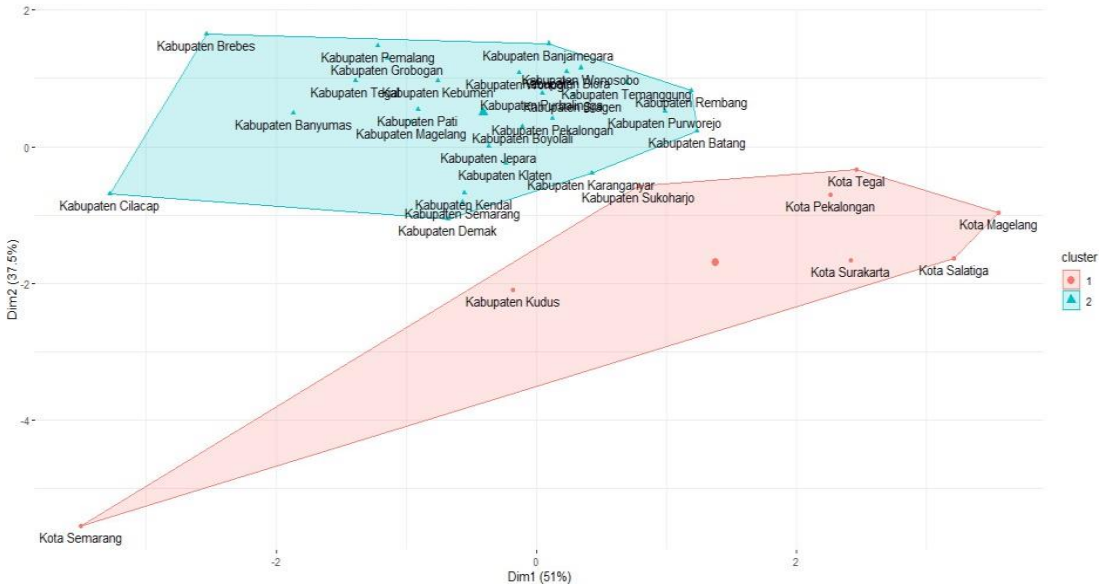

Figure 3. Clustering Results Plot Using the K-Means Method

Based on the clustering of regencies/cities shown in Figure 3, two clusters of regencies/cities were formed. The first cluster consisted of 8 members, while the second cluster consisted of 27 members, as shown in Table 8.

Table 8. Clustering Results Using the K-Means Method

Cluster	Regencies/Cities	Results Identification
1	Semarang City, Surakarta City, Magelang City, Salatiga City, Pekalongan City, Tegal City, Kudus Regency, and Sukoharjo Regency.	This cluster has a lower population and labor force compared to Cluster 2, but a higher average Regency/City Minimum Wage (UMK). These areas reflect urban regions with better employment opportunities, resulting in a more controlled unemployment rate.
2	Cilacap, Banjarnegara, Purbalingga, Pemalang, Tegal, Brebes, Batang, Pekalongan, Kendal, Temanggung, Semarang, Demak, Jepara, Kudus, Pati, Blora, Grobogan, Rembang, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Purworejo, Kebumen, Banyumas, and Wonosobo Regencies.	The high population and labor force in these areas potentially increase the challenges in managing the open unemployment rate.

After the clustering process is complete, the next step is to perform a characteristic analysis for each cluster to determine the average value of each variable. The characteristics of each cluster are presented in Table 9.

Table 9. Cluster Characteristics Based on Mean Values

Variable	Cluster 1	Cluster 2
Total Population	512901	1176393
Total Minimum Weight (UMK)	2326700	2160596
Total Workforce	339793	679659

After clustering based on the open unemployment rate in 2023 in the regencies/cities of Central Java Province, the characteristic analysis showed that Cluster 1, which included 8 regencies/cities, had an average population of 512.901 people, an average regency/city Minimum Wage (UMK) of IDR 2.326.700, and an average labor force of 339.793 people. Meanwhile, Cluster 2, which included 27 regencies/cities, had an average population of 1.176.393 people, an average UMK of IDR 2.160.596, and an average labor force of 679.659 people. These data indicate significant differences in characteristics between Cluster 1 and Cluster 2, particularly in terms of population, labor force, and UMK levels. These differences can serve as a basis for formulating more focused and tailored policies for each cluster, especially in managing the open unemployment rate in Central Java Province. However, in addition to using the Silhouette Coefficient as a measure of cluster quality, this analysis also considered the consistency of the results by comparing the hierarchical clustering and K-Means methods. The results obtained showed relatively similar clustering patterns, indicating that the clustering results were quite stable and insensitive to differences in the methods used.

Comparison between the Average Linkage and K-Means Clustering Methods

The assessment of the two clustering methods, namely K-Means and Average Linkage, was conducted using the Silhouette Coefficient to measure the quality of the clustering results. The evaluation of clustering quality showed that the optimal cluster for the Average Linkage method is k = 3, while for the K-Means Clustering method, the optimal cluster is k = 2. The Silhouette Coefficient values are presented in Table 10.

Table 10. Clustering Quality Evaluation Results Using the Silhouette Coefficient

Metode	Silhouette Coefficient
Average Linkage	0.444
K-Means	0.676

Based on the comparison of clustering result using the Silhouette Coefficient, it is evident that the K-Means Clustering method produces a higher value of 0.676 compared to the Average Linkage method with a value of 0.444. This indicate that the cluster structure generated by K-Means is better and has clearer separation. A silhouette coefficient value of 0.676 indicates that the resulting cluster quality falls into the good category, whereas the value of 0.444 for the Average Linkage method falls into the moderate category. This difference suggests that the K-Means method is more optimal in grouping the data compared to the Average Linkage method.

Discussion

Based on the analysis results, the K-Means Method yielded more optimal Clustering than the Average Linkage method. This is indicated by the higher Silhouette Coefficient value obtained by K-Means. The Silhouette Coefficient value of 0.676 for K-Means falls into the good category, while the Average Linkage method only reaches 0.444, which indicates a moderate cluster structure. This result indicates that K-Means produces more compact cluster with clearer separation compared to Average Linkage, as it minimizes within cluster variance. In contrast, Average Linkage relies on average distances, which may lead to less distinct cluster boundaries. This study aligns with the research by (Arifatul Ulya et al., 2023), which showed that the K-Means Clustering method is effective in grouping regencies/cities based on the open unemployment rate indicators in Central Java Province. In their study, K-Means successfully formed 4 clusters with an accuracy of 66.8%, supporting the findings of this study. In addition, the Average Linkage method produced three cluster with one single member cluster, indicating possible over segmentation, while K-Means produced two more balanced clusters that are easier to interpret. The analysis in this study showed that the K-Means Clustering method with 2 clusters is the most optimal approach for grouping regions based on population, Regency/City Minimum Wage (UMK), and labor force, with a Silhouette Coefficient value of 0.676, which is higher than the Average Linkage method that only reached 0.444. This clustering divides Central Java Province into two clusters: the first cluster consists of 8 regencies/cities, and the second cluster consists of 27 regencies/cities. The consistency of these results reinforces that the K-Means

Clustering method is an appropriate approach for classifying regencies/cities based on the open unemployment rate in Central Java Province and can be used as a basis for formulating more targeted unemployment mitigation policies.

Conclusion

Based on the analysis results, three groups of regencies/cities were formed using the Average Linkage method, and two groups were formed using the K-Means Clustering method. From the evaluation using the Silhouette Coefficient, it can be concluded that the K-Means method provides more optimal clustering results compared to the Average Linkage method in grouping regencies/cities based on the open unemployment rate in Central Java Province in 2023, with a value of 0.676 In the K-Means Clustering method, the first group consists of 8 regencies/cities. This group has a lower population and labor force compared to Cluster 2, but a higher average regency/city Minimum Wage (UMK). These areas reflect urban regions with better employment opportunities, resulting in a more controlled open unemployment rate. However, this study has several limitations, including its sensitivity to outliers and the use of limited variables. Therefore, future research is recommended to incorporate additional relevant factors and apply more robust methods to improve the reliability of the clustering results.

References

Abdulhafedh, A. (2025). Applying K-Means Clustering and Fuzzy C-Means Clustering in Vehicle Crashes. *Open Access Library Journal*, *12*, 1–11. https://doi.org/https://doi.org/10.4236/oalib.1112856

Arifatul Ulya, F., Abdullah, A. N., Aisya Hanan, T., & Manfaati Nur, I. (2023). Pengelompokkan Tingkat Pengangguran Terbuka Di Jawa Tengah Menggunakan Metode K – Means Clustering. *Journal Of Data Insights*, *1*(2), 71–80. https://doi.org/10.26714/jodi.v1i2.148

Azhar, Z. A., Handajani, S. S., & Slamet, I. (2024). Pemodelan Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran Terbuka (TPT) Di Provinsi Jawa Tengah Menggunakan Regresi Spline Truncated Multivariabel. *Jurnal Sutasoma*, *2*(2), 66–73. https://doi.org/10.58878/sutasoma.v2i2.264

Azzahra, L., & Amru Yasir. (2024). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *3*(1), 1–10. https://doi.org/10.70340/jirsi.v3i1.88

Butar Butar, R. P. (2023). Analisis Hierarchical Dan Non-Hierarchical Clustering Untuk Pengelompokkan Potensi Ekonomi Kelautan Indonesia 2021. *Jurnal*

Sistem Dan Teknologi Informasi (JustIN), *11*(3), 543. https://doi.org/10.26418/justin.v11i3.67283

Chan, J. Y. Le, Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Full Collinearity approach is applied because it identifies collinearity across all variables in the model, irrespective of where they are placed in the model. *Mathematics*, *10*(8).

Doni, A. H., Alfiona, F., Andespa, W., & Al-Amin, A.-A. (2023). Pengangguran Dalam Perspektif Ekonomi Islam Dan Kovensional. *Jurnal Ilmiah Ekonomi, Manajemen Dan Syariah*, *2*(3), 1–10. https://doi.org/10.55883/jiemas.v2i3.20

Frisnoiry, S., Sihotang, H. M. W., Indri, N., & Munthe, T. (2024). Analisis Permasalahan Pengangguran Di Indonesia. *Jurnal Ilmiah Komputerisasi Akuntansi*, *17*(1). https://doi.org/https://doi.org/10.51903/kompak.v17i1.1866

Hasmira, Wahidah Alwi, & Khalilah Nurfadilah. (2023). Penentuan Cluster Hirarki Optimum dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan dengan Indikator Kemiskinan. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, *11*(1), 93–102. https://doi.org/10.24252/msa.v11i1.33910

Hegelund, E., & Taalbi, J. (2023). What determines unemployment in the long run? Band spectrum regression on ten countries 1913–2016. *Structural Change and Economic Dynamics*, *64*(April 2022), 144–167. https://doi.org/10.1016/j.strueco.2022.11.009

Iqbal, M., & Nurul Huda, M. (2023). Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Mengelompokkan Daerah Penyebaran COVID-19 di Kabupaten Bogor. *Jurnal Ilmiah Matematika*, *2*(1), 47–56. https://doi.org/https://doi.org/10.30872/dasar.v2i1.1019

Kurniawan, A. (2024). Determinan Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah : Analisis Regresi Data Panel. *Jurnal Litbang Provinsi Jawa Tengah*, *22*(2), 123–136. https://doi.org/https://doi.org/10.36762/jurnaljateng.v22i2.1217 123

Laber, E. S., & Batista, M. (2024). On the cohesion and separability of average-link for hierarchical agglomerative clustering. *Advances in Neural Information Processing Systems*, *37*(NeurIPS).

Maisuun, N. N. F., & Darmawanti, I. (2023). Gambaran Strategi Coping Pada Pengangguran Perempuan Dewasa Awal. *Character: Jurnal Penelitian Psikologi*, *10*(02), 80–102. https://doi.org/https://doi.org/10.26740/cjpp.v10i2.53388

Messakh, G. C., Hayati, M. N., & Sifriyani, S. (2023). Penerapan Metode K-Means Dalam Pengelompokan Kabupaten/Kota Di Kalimantan Berdasarkan Indikator Pendidikan. *Eksponensial*, *14*(2), 57. https://doi.org/10.30872/eksponensial.v14i2.1103

Nugroho, D. W., Bramhatchi, F., Wulandari, S. P., & Eka, A. (2024). Pengelompokan Indikator Kesejahteraan Masyarakat Berdasarkan Kabupaten/Kota di Jawa Tengah Tahun 2023 Menggunakan Analisis Cluster. *Switch: Jurnal Sains Dan*

Teknologi Informasi, *2*, 87–101. https://doi.org/https://doi.org/10.62951/switch.v2i5.285

Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *6*(2), 48. https://doi.org/10.51557/pt_jiit.v6i2.659

Rahmi, M. A., Arum, P. R., & Wahyu Utami, T. (2025). Implementation of K-Means Algorithm to Group Age of Cardiovascular Disease Patients. *Journal Of Data Insights*, *3*(1), 42–47. https://doi.org/10.26714/jodi.v3i1.216

Setyaningsih, L. I., Wulandari, A. R., & Arum, P. R. (2023). Implementation of K-Means to Classify Poverty Based on Housing Characteristics in Central Java in 2021. *Jurnal Pendidikan Matematika (Kudus)*, *6*(1), 49. https://doi.org/10.21043/jpmk.v6i1.19529

Sinta Amelia, A. M. (2022). *Berdasarkan Indikator Indeks Pembangunan Manusia*. *04*(02), 69–74.

Syalomitha R, S., Nadhira F, S., Anistya A, G., Raihana Z, R., Dominique P S, J., Aliyyah N, C., Putri P, A., & Toharudin, T. (2021). Pengelompokkan kabupaten/kota di Provinsi Jawa Barat menurut indikator indeks pembangunan manusia tahun 2020. *E-Prosiding Seminar Nasional Statistika| Departemen Statistika FMIPA Universitas Padjadjaran*, *10*, 108–123. https://doi.org/https://doi.org/10.1234/pns.v10i.80

Xu, N., Finkelman, R. B., Dai, S., Xu, C., & Peng, M. (2021). Average Linkage Hierarchical Clustering Algorithm for Determining the Relationships between Elements in Coal. *ACS Omega*, *6*(9), 6206–6217. https://doi.org/10.1021/acsomega.0c05758

Yusniyanti, A. L., Virgantari, F., & Faridhan, Y. E. (2021). Comparison of Average Linkage and K-Means Methods in Clustering Indonesia's Provinces Based on Welfare Indicators. *Journal of Physics: Conference Series*, *1863*(1). https://doi.org/10.1088/1742-6596/1863/1/012071

Zubair, M., Iqbal, M. A., Shil, A., Chowdhury, M. J. M., Moni, M. A., & Sarker, I. H. (2022). An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling. *Annals of Data Science*. https://doi.org/10.1007/s40745-022-00428-2