PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 21, No. 1 March 2025 | DOI: 10.33480/pilar.v21i1.6445

AGGLOMERATIVE HIERARCHICAL CLUSTERING FOR REGIONAL GROUPING IN CENTRAL JAVA BASED ON WELFARE INDICES

Raafi Kurnia Desita[1]; Amiq Fahmi[2*]; Asih Rohmani[3]; MY. Teguh Sulistyono[4]

Sistem Informasi[1,2,3,4]
Universitas Dian Nuswantoro, Kota Semarang, Indonesia[1,2,3,4]
https://dinus.ac.id[1,2,3,4]
112202106599@mhs.dinus.ac.id[1], amiq.fahmi@dsn.dinus.ac.id[2*], aseharsoyo@dsn.dinus.ac.id[3], teguh.sulistyono@dsn.dinus.ac.id[4]
(*) Corresponding Author

Abstract— *Central Java Province comprises 35 regencies/cities with diverse welfare characteristics. These variations present challenges for the government in formulating targeted development policies. This study aims to group regions in Central Java based on welfare indices to support more effective policy planning. The Agglomerative Hierarchical Clustering method with the Average Linkage approach is applied to cluster the regions based on three attributes: Human Development Index, Uninhabitable Houses, and Economic Growth Rate. Data were obtained from the Central Java Provincial Social Service and the official website of the Central Statistics Agency (BPS) and processed using the proposed method. Experimental results indicate three clusters with proportions: 32 regions in cluster 1 (91.4%), 2 regions in cluster 2 (5.7%), and 1 region in cluster 3 (2.9%). Regions with higher welfare dominate the first cluster, while the second and third clusters include regions facing more significant welfare challenges. Clustering results were evaluated using the Silhouette Score (0.542) and Davies-Bouldin Index Score (0.613), demonstrating that the applied method effectively grouped regions based on the specified attributes. The findings of this study are anticipated to lay the groundwork for more directed and effective development policies.*

Keywords: *agglomerative hierarchical clustering, average linkage, central java, welfare indices.*

Abstrak—*Provinsi Jawa Tengah terdiri dari 35 kabupaten/kota dengan karakteristik kesejahteraan masyarakat yang beragam. Perbedaan ini menjadi tantangan bagi pemerintah dalam merumuskan kebijakan pembangunan yang tepat sasaran. Penelitian ini bertujuan mengelompokkan wilayah di Jawa Tengah berdasarkan indeks kesejahteraan masyarakat guna mendukung perencanaan kebijakan yang lebih efektif. Metode Agglomerative Hierarchical Clustering dengan pendekatan Average Linkage diterapkan untuk mengelompokkan wilayah berdasarkan tiga atribut: Indeks Pembangunan Manusia, Rumah Tidak Layak Huni, dan Laju Pertumbuhan Ekonomi. Data diperoleh dari Dinas Sosial Provinsi Jawa Tengah serta situs resmi Badan Pusat Statistik (BPS), dan diolah menggunakan metode yang diusulkan. Hasil eksperimen menunjukkan tiga klaster dengan proporsi: 32 wilayah dalam klaster 1 (91,4%), 2 wilayah dalam klaster 2 (5,7%), dan 1 wilayah dalam klaster 3 (2,9%). Klaster pertama didominasi oleh wilayah dengan kesejahteraan yang lebih tinggi, sementara klaster kedua dan ketiga mencakup wilayah dengan tantangan kesejahteraan yang lebih signifikan. Hasil klasterisasi dievaluasi menggunakan Silhouette Score (0,542) dan Davies-Bouldin Index Score (0,613), menunjukkan metode yang diterapkan berhasil mengelompokkan wilayah dengan baik berdasarkan atribut yang ditetapkan. Temuan penelitian ini diharapkan menjadi dasar dalam kebijakan pembangunan yang lebih terarah dan efektif.*

Kata Kunci*: pengelompokan hierarkis aglomeratif, keterkaitan rata-rata, jawa tengah, kesejahteraan masyarakat.*

INTRODUCTION

Public welfare is one of the goals of the Indonesian state, as stated in paragraph IV of the Preamble to the Undang-Undang Dasar 1945 Constitution. However, to date, one of the main challenges in realizing this welfare is the high poverty rate, which continues to increase in Indonesia (Agustina & Yahya, 2022). Poverty causes a decline in the quality of human resources, low productivity, and limited access to education and health. This condition is a significant obstacle to development and progress in various provinces, regencies, and cities in Indonesia (Agustina & Yahya, 2022; Dalimunthe, 2021).

Various indicators, such as population density, open unemployment rate, life expectancy, Human Development Index, and the percentage of poor people, can be used to evaluate community welfare in an Indonesian region (Rahmawati et al., 2022). Diverse studies show that the Human Development Index (HDI) significantly negatively influences poverty levels (Wardani et al., 2023).

Central Java Province has Indonesia's fifth highest population density and will be the second poorest province in Java in 2023 (Asyfani et al., 2024; Larasati et al., 2024). This province has an area of 3.43 million hectares and is divided into 29 districts and 6 cities with a total population of 37,892,280 people. Of this number, 3,704,330 people (10.47%) are categorized as poor (Provinsi Jawa Tengah, 2024). (Provinsi Jawa Tengah, 2024). Although economic growth and the human development index have increased, poverty rates have not decreased significantly (Fuady et al., 2022).

The diverse welfare conditions of the community in each regency/city in Central Java show significant gaps, creating various challenges that require a tailored approach to overcome them. Although the allocation of village funds has shown the potential to reduce poverty levels, its direct impact on the welfare of the community as a whole is still limited (Hilmawan et al., 2023).

Today, information technology, especially in data analysis and artificial intelligence, has played an important role in providing data-based insights for policymaking. One practical approach that offers innovative solutions to address challenges in public welfare issues is data mining and machine learning methods, such as clustering analysis.

Clustering is an unsupervised learning technique that groups data based on similarities in specific properties or characteristics while maximizing the dissimilarity between clusters (Pradana & Ha, 2021; Thakare & Sonawane, 2023). Clustering techniques are divided into two types: hierarchical and nonhierarchical. Hierarchical clustering is a data mining technique that creates a tree-like cluster structure by measuring the proximity distances between objects. It is known as a dendrogram, which visually represents the relationships between groups. This technique has two clustering strategies: divisive (top-down) and agglomerative (bottom-up) (Muflihan et al., 2022; Repiská et al., 2022; Sharma, 2022). The Agglomerative Hierarchical Clustering method has better advantages because it does not require determining the number of clusters initially and can display visual results as a dendrogram.

To overcome the inequality of population distribution and poverty in Central Java Province, which impacts unemployment, crime, and low productivity, Asyfani et al. (Asyfani et al., 2024) used the Hierarchical Clustering Ward method with optimal cluster determination using the elbow method. Visualizing the results through a dendrogram produces four regency/city clusters, which provide strategic information for data-based policies. However, this study only focuses on a single demographic aspect without considering other factors contributing to community welfare.

In a broader context, it is important to consider the various factors that affect societal well-being. These factors include life expectancy, quality of education, income, the percentage of uninhabitable houses, and the level of economic growth over a certain period of time. Research that combines these various aspects will provide a more comprehensive picture of a region's socioeconomic conditions.

Larasati et al. (Larasati et al., 2024) focused on the high poverty and unemployment rates that impact crime and low productivity. They identified regency/city that requires priority handling using poverty indicators such as the Poverty Line, Percentage of Poor Population, Poverty Severity Index, Poverty Depth Index, Distribution of Expenditure Based on World Bank Criteria, Number of Unemployed, Average Years of Schooling, and Expected Years of Schooling. The clustering method uses Agglomerative Hierarchical Clustering (AHC) with single linkage, complete linkage, and average linkage. The analysis results show two poverty levels: cluster 1 with high poverty (31 regencies/cities) and cluster 2 with low poverty (4 cities). The government is recommended to improve social welfare, education, and employment programs in Cluster 1. However, this study did not involve the Human Development Index (HDI), Uninhabitable Houses, and Economic Growth Rate.

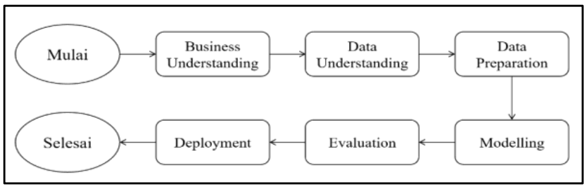
This study aims to group regions in Central Java based on the community welfare index to support more effective policy planning. This study overcomes the limitations of previous studies (Asyfani et al., 2024; Larasati et al., 2024) using the Agglomerative Hierarchical Clustering (AHC) method with the Average Linkage approach. This analysis considers various social and economic aspects, such as the HDI, which includes life expectancy, quality of education and income, the percentage of uninhabitable houses, and a region's economic growth level in a certain period. The results of this study are expected to provide strategic contributions in formulating more

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 21, No. 1 March 2025 | DOI: 10.33480/pilar.v21i1.6445

effective policies to improve community welfare in Central Java Province.

MATERIALS AND METHODS

This study methodology uses a quantitative approach to group regions based on similar characteristics in indicators of community welfare. This method includes a series of stages in the Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM is a powerful method and framework that effectively manages data mining, especially in comprehensively solving quantitative research problems (Tripathi et al., 2021). The stages of the CRISP-DM method can be glimpsed in Figure 1.



Source: (Research Result, 2025)
Figure 1. CRISP-DM Method

Business Understanding

The business understanding stage aims to understand the context of the problem to be solved. This study focuses on mapping the regencies/cities in Central Java Province based on the community welfare index, explicitly using the Human Development Index (HDI), Uninhabitable Houses, and Economic Growth Rate. This approach is intended to support formulating more effective and targeted policies.

Data Understanding

The data used in this study were sourced from the Central Java Provincial Social Service and the Badan Pusat Statistik (BPS) official website at the city and regency levels in the province. The research attributes include various indicators of community welfare, including Regency/City Name, Human Development Index (HDI), Number of Uninhabitable Houses, and Economic Growth Rate in each regency/city in Central Java Province. These indicators were chosen because they can provide a comprehensive picture of poverty's social and economic aspects. At this stage, initial data exploration was carried out to understand the characteristics of the data, such as statistical descriptions, data distribution, and correlations between attributes.

Data Preparation

Data preparation is critical for effective cluster analysis as it significantly impacts the quality of the clustering results. This process involves several key steps, including data cleaning, removing irrelevant or incorrect data, normalization to ensure features are on the same scale, and feature selection to identify the most relevant attributes for clustering (Olexandr TKACHYK, 2023).

The tool used in this study is Python, where the libraries used include Pandas, Numpy, Matplotlib, Sklearn, and Scipy. In this dataset, no missing values or duplicate data were found, so the data preparation stage was continued with data normalization or aligning the variable scale using the StandardScaler method (Zulkifilu & Yasir, 2022). StandardScaler helps clustering by ensuring that all features have similar scales so that no single feature dominates, leading to better clustering. Eq. (1) shows scaling using Z-Score.

$$z = \frac{x-\mu}{\sigma} \quad \text{......} (1)$$

Where is
z = z-score value or fundamental value
x = actual data value
μ = average of all data
σ = standard deviation of all data

Modelling

After successfully preparing the data, the next step is to perform clustering. This study uses clustering to group the data into specific clusters based on similar characteristics (Dinata & Syaputra, 2020) Hierarchical clustering was chosen because of its advantages, such as flexibility, dendrogram visualization, no need for labelling, handling data at various scales, and ease of interpretation. It is used as an approach to generate cluster hierarchies (Boyko & Tkachyk, 2023).

In this study, Agglomerative Hierarchical Clustering was applied to group similar data gradually, starting from the smallest group to forming larger groups (bottom-up) (Sharma, 2022; Thakare & Sonawane, 2023). Distance measurements between data are performed using Euclidean Distance, which measures the linear distance between two points in multidimensional space (Karmanita et al., 2024). The linkage criteria used is Average Linkage, where the distance between two clusters is defined based on the average distance between all pairs of points from two different clusters. This stage will produce a dendrogram to illustrate the hierarchical clustering of regions.

Evaluation

At this stage, the clustering results will be evaluated to assess the quality of the clusters. This study's evaluation includes the Silhouette Score and the Davies-Bouldin Index.

1. Silhouette score
Silhouette Score measures the accuracy of each data point's location in its cluster. The higher the Silhouette Score value, the better the quality of the resulting cluster (Hasan, 2024). Table 1 shows the cluster quality assessment using the Silhouette Score.

Table 1 Silhouette Score Range

Silhouette Score (SS) Value	Information
0,7 < SS ≤ 1	Strong Structure
0,5 < SS ≤ 0,7	Moderate Structure
0,25 < SS ≤ 0,5	Weak Structure
SS ≤ 0,2	No Structure

Source : (Nikum & Yuliansyah, 2024)

2. Davies-Bouldin Index (DBI) Score
Davies-Bouldin Index (DBI) Score is an evaluation matrix that evaluates how well the separation between various clusters occurs. The lower the DBI Score, the better the cluster separation (Hasan, 2024).

Deployment

The results of this study will be visualized in the form of graphs and interpretations of cluster characteristics. This graph will show cluster information for each district/city in Central Java Province. These results can be used as a reference for local governments to identify regional development priorities and formulate more targeted development policies.

RESULTS AND DISCUSSION

This study uses secondary data from 35 districts/cities in Central Java Province. including Banjarnegara, Banyumas, Batang, Blora, Boyolali, Brebes, Cilacap, Demak, Grobogan, Jepara, Karanganyar, Kebumen, Kendal, Klaten, Magelang City, Pekalongan City, Salatiga City, Surakarta City Semarang City, Tegal City, Kudus, Magelang, Pati, Pekalongan, Pemalang, Purbalingga, Purwokerto, Rembang, Semarang, Sragen, Sukoharjo, Tegal, Temanggung, Wonogiri, and Wonosobo. The data includes three attributes: Human Development Index (HDI), Uninhabitable Houses, and Economic Growth Rate. Table 2 shows the dataset used in the clustering process.

Table 2. Ready to Use Dataset

Regency/ City	HDI	Uninhabitable Houses	Economy Growth Rate
Banjarneg ara	69,16	14558	4,98
Banyumas	73,86	13941	5,40
Batang	70,20	11910	5,53
Blora	70,63	43212	3,10
Boyolali	75.41	5283	5,63
...	...	...	...
Sukoharjo	78,65	159	5,06
Tegal	71,12	1850	4,93
Temanggu ng	71,33	9335	5,00
Wonogiri	71,97	5449	4,98
Wonosob o	70,18	18594	4,30

Source: (Research Result, 2025)

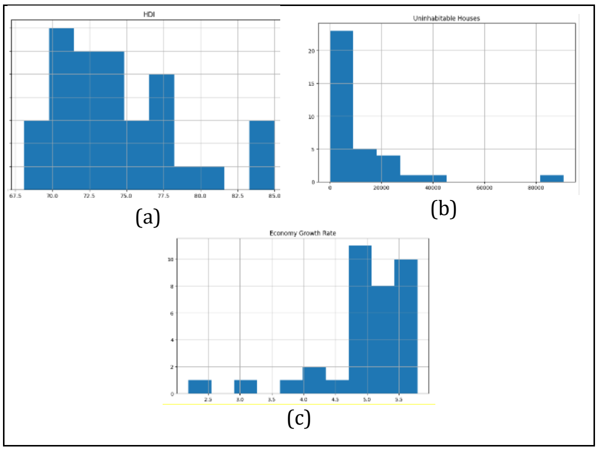
A descriptive analysis was conducted on the selected attributes to gain deeper insights into the dataset. The statistical analysis results in Table 3 provide an overview of data distribution, including the mean, median, minimum, and maximum values for each attribute.

Table 3. Statistical Description

	HDI	Uninhabitable Houses	Economy Growth Rate
Count	35,000000	35,000000	35,000000
Mean	74, 302286	10654,600000	4,985143
Std	4, 282907	17217,447126	0,741931
Min	68,080000	6,000000	2,190000
25%	71,390000	924,500000	4,980000
50%	73,850000	5268,000000	5,070000
75%	76,710000	12925,500000	5,445000
max	84,990000	90679,000000	5,790000

Source: (Research Result, 2025)

In addition to statistical description analysis, visualizing data distribution using bar charts helps illustrate the distribution pattern more clearly. Figure 2 shows the histogram of the data distribution of the third attribute.



Source: (Research Result, 2025)
Figure 2 (a) Histogram of HDI Data Distribution (b) Histogram of Uninhabitable Houses Data Distribution (c) Histogram of Economic Growth Rate Data Distribution

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 21, No. 1 March 2025 | DOI: 10.33480/pilar.v21i1.6445

Data Preparation

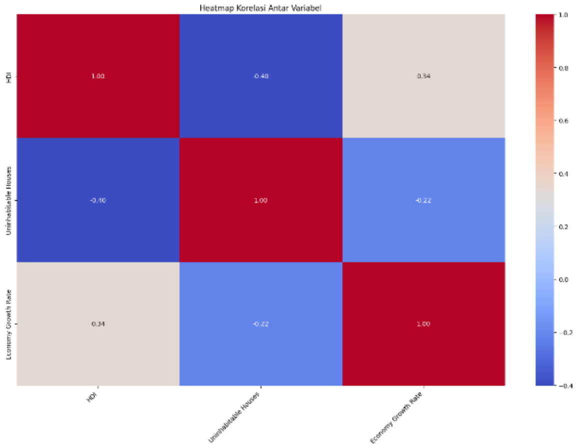
At this stage, the data will be normalized using the Standard Scaler method. This method changes the distribution of data by subtracting the mean and dividing by the standard deviation. Thus, all features (attributes) in the dataset will have the same scale. The data used has undergone a normalization process to equalize the scale of each feature. The normalization results are shown in Table 4. After normalization, the values of each attribute have a distribution with a mean close to zero and a standard deviation of one. For example, the Human Development Index (HDI) value in Banjarnegara is -1.218, while in Sukoharjo, it is 1.029, indicating a significant difference between regions with low and high HDI. A similar pattern is observed in other attributes, where regions with higher economic growth rates have larger normalized values compared to those with lower growth rates.

Table 4. Dataset after Normalization

Regency/ City	HDI	Uninhabitable Houses	Economy Growth Rate
Banjarnegara	-1,21818177	0,23002173	-0,00703292
Banyumas	-0,10477527	0,19366281	0,5673223
Batang	-0,97181097	0,07397891	0,74509892
Blora	-0,86994612	1,9185606	-2,57795629
Boyolali	0,26241197	-0,31654064	0,88185016
...	...	...	...
Sukoharjo	1,02995177	-0,61849056	0,10236807
Tegal	-0,75386757	-0,51884237	-0,07540854
Temanggung	-0,70411962	-0,07776212	0,02031733
Wonogiri	-0,55250682	-0,3067585	-0,00703292
Wonosobo	-0,97654887	0,46785738	-0,93694137

Source: (Research Result, 2025)

To comprehend the interrelationship among attributes, a correlation analysis was conducted utilizing the Correlation Matrix. The Correlation Matrix is a statistical technique employed to assess the association between two variables, with correlation values ranging from -1 to 1. Each cell within this matrix contains the correlation coefficient of a specific pair of variables, illustrating the strength and direction of their relationship. Figure 3 presents the outcomes of the Correlation Matrix for the attributes within the dataset.



Source: (Research Result, 2025)

Figure 3. Correlation Matrix between Attributes

Based on the correlation results, HDI has a positive correlation with Economy Growth Rate (0.34), indicating that regions with higher economic growth tend to have better HDI levels. Conversely, the number of Uninhabitable Houses has a negative correlation with HDI (-0.10) and Economy Growth Rate (-0.22), suggesting that areas with more uninhabitable houses tend to have lower HDI and slower economic growth. These findings highlight the role of economic growth in improving HDI, while poor housing conditions are linked to lower quality of life and slower regional economic development.

Modelling

The next stage in this study is the modelling stage, where the clustering method used is Agglomerative Hierarchical Clustering (AHC). Table 5 shows the best combination of parameters for building a hierarchical cluster model. Based on the data in the table, the best approach is Average Linkage. In this process, the distance between data will be calculated using Euclidean Distance, one of the standard matrices used to measure the distance between points in multidimensional space.

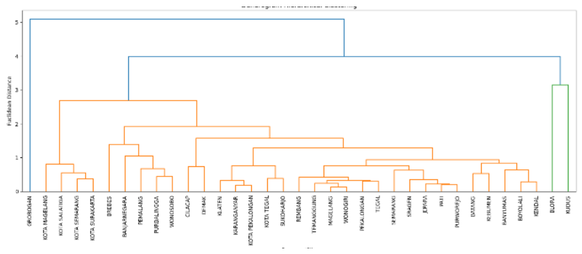
Table 5. Best Parameter Combination

Cluster	Parameter	Silhouette Score
3	Average Linkage Euclidean Distance	0,542
4	Average Linkage Euclidean Distance	0,485
5	Complete Linkage Euclidean Distance	0,427

Source: (Research Result, 2025)

Using this AHC method, the data will be clustered gradually from N clusters into one cluster.

The clustering results can be visualized in the dendrogram graph in Figure 4.



Source: (Research Result, 2025)
Figure 4. Clustering Dendogram

Evaluation

Cluster evaluation was conducted using the Silhouette Score and Davies-Bouldin Score. Silhouette Score is one of the techniques used to measure how well an object is located in a cluster, which can validate either one data group or all data groups at once (Hasan, 2024; Nikum & Yuliansyah, 2024). Silhouette Score has a value range between -1 to 1, where a value close to 1 indicates that the clustering has good quality. Generally, a value above 0.5 is considered good enough or adequate. In addition to the Silhouette Score, the Davies-Bouldin Index (DBI) Score is also used as an evaluation metric to assess the quality of clustering results. The smaller the DBI value obtained (non-negative), the better the cluster produced by the clustering algorithm (Hasan, 2024).

Table 6 shows the Silhouette Score and Davies-Bouldin Index (DBI) Score values.

Table 6. Cluster Evaluation

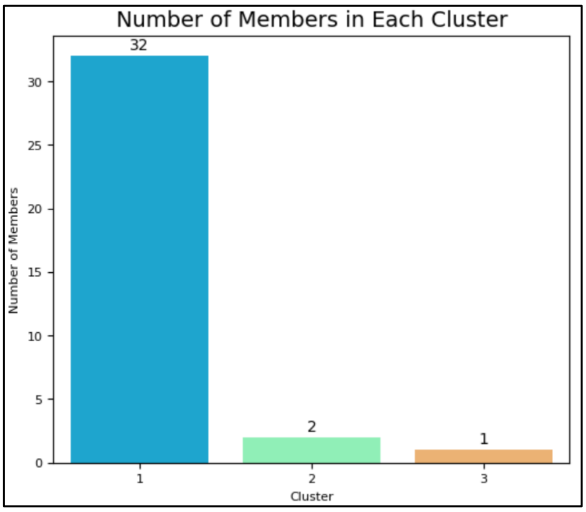
Evaluation Matrix	Score
Silhouette Score	0,54189
Davies-Bouldin Score	0,61277

Source: (Research Result, 2025)

Based on the evaluation results using the Silhouette Score and Davies-Bouldin Index (DBI) Score, the value 0,542 indicates that the objects have relatively high similarity and are well separated from other clusters. The DBI value of 0,613 indicates that, in general, the clusters formed have quite a large density and distance.

Deployment

After cutting the hierarchical tree into three clusters, three data groups were obtained: cluster 1 consists of 32 regencies/cities, cluster 2 consists of 2 regencies, and cluster 3 consists of 1 regency. Figure 5 shows the frequency of members of each cluster.



Source: (Research Result, 2025)
Figure 5. Cluster Frequency Diagram

Table 7 shows the distribution of cluster members for the Central Java Province, obtained using the Average Linkage method. This method groups data based on the average distance between all pairs of points from two different clusters. The first cluster consists of 32 regencies/cities, the second of 2 regencies, and the third of 1 regency, showing a different distribution of characteristics in each region.

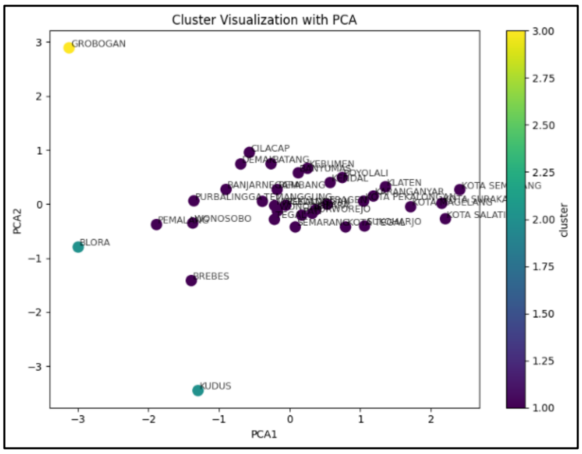
Table 7. Cluster Distribution

Cluster	Total Members	Proportion	Cluster Members
1	32	0,914286	Banjarnegara, Banyumas, Batang, Brebes, Boyolali, Cilacap, Demak, Jepara, Karanganyar, Kebumen, Kendal, Klaten, Kota Magelang, Kota Pekalongan, Kota Salatiga, Kota Semarang, Kota Surakarta, Kota Tegal, Magelang, Pati, Pekalongan, Pemalang, Purbalingga, Purworejo, Rembang, Semarang, Sragen, Sukoharjo, Tegal, Temanggung, Wonogiri Wonosobo
2	2	0,057143	Blora, Kudus
3	1	0,028571	Grobogan

Source: (Research Result, 2025)

From the clustering results, Principal Component Analysis (PCA) will be used to visualize the distribution of the formed clusters. Figure 6 shows the distribution of clusters, where purple represents cluster 1, blue represents cluster 2, and yellow represents cluster 3.

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 21, No. 1 March 2025 | DOI: 10.33480/pilar.v21i1.6445



Source: (Research Result, 2025)
Figure 6. Visualization of Cluster Distribution with PCA

Table 8 shows the statistical description for the three regional clusters, which includes various important indicators such as the mean value, standard deviation, and amount of data for each attribute in each cluster.

Table 8. Statistical Description of Each Cluster

		Cluster 1	Cluster 2	Cluster 3
HDI	Mean	74,429688	73,67	71,49
	Std	4,384580	4,299209	NaN
	Count	32	2	1
Uninhabitable Houses	Mean	7451,46875	21892,5	90679,0
	Std	8395,711974	30150,326043	NaN
	Count	32	2	1
Economy Growth Rate	Mean	5,131563	2,645	4,98
	Std	0,463612	0,643467	NaN
	Count	32	2	1

Source: (Research Result, 2025)

1. Cluster 1

Cluster 1 is a group consisting of 32 regions. The average Human Development Index in this cluster is 74, 43, indicating that the regions in this cluster have a relatively high level of human development. In terms of economic growth, cluster 1 is quite stable, with an average value of 5, 13%, reflecting good economic conditions. However, the average number of Uninhabitable Houses in cluster 1 is quite high, reaching 7.451.

2. Cluster 2

Cluster 2 consists of 2 regions. In this cluster, the average Human Development Index value is 73, 67, slightly lower than in cluster 1. The economic growth rate in this cluster is relatively low compared to other clusters, at an average of 2, 65%. Meanwhile, the number of Uninhabitable Houses in this cluster reaches 21.892. This reflects a more serious economic and housing problem than in Cluster 1.

3. Cluster 3

Cluster 3 consists of only 1 region, with the lowest average Human Development Index value, 71, 49. Although its economic growth rate is close to that of cluster 1, 4, 98%, the average number of Uninhabitable Houses in this cluster is very high, reaching 90.679. This reflects that, beyond their good economic growth potential, regions in this cluster have a low level of human development and housing problems.

Based on these findings, the results of this study can be used as a reference for the Jawa Tengah provincial government in designing development policies. Areas in cluster 3 need more attention with increased housing improvements, enhancing basic infrastructure, increasing access to education and healthcare services, and implementing economic empowerment programs to improve community welfare. For cluster 2, the government must encourage economic growth by attracting investment in the industrial and trade sectors, building job training centers, and repairing uninhabitable housing. Meanwhile, cluster 1 still needs to be supported to continue to develop through a program to renovate uninhabitable houses and increase investment in infrastructure and the creative industry.

CONCLUSION

Based on the discussion conducted in this study, it can be concluded that grouping regions in Central Java Province using the Agglomerative Hierarchical Clustering (AHC) method with the Average Linkage approach produces three clusters. Cluster 1 reflects a developed region with a high Human Development Index (HDI), stable economic growth, and better infrastructure, although it still faces inequality in the number of Uninhabitable Houses. Cluster 2 has a lower HDI, limited economic growth, and a higher number of Uninhabitable Houses. Cluster 3 reflects the most disadvantaged region, with the lowest HDI and a very high number of Uninhabitable Houses, although it has good economic growth. Therefore, cluster 3 needs to be the main priority in development policies, because it has the most disadvantaged conditions compared to other clusters.

This study can be further developed by exploring other clustering methods, such as K-Means, DBSCAN, or other algorithms, to compare the results obtained. In addition, integration with predictive methods such as regression or machine learning can be developed to help predict future trends in community welfare so that the policies formulated can be more proactive and sustainable.

REFERENCEE

Agustina, R. E., & Yahya, A. (2022). Pembangunan Desa Dan Tingkat Kemiskinan Kabupaten Di Indonesia. *Jurnal Litbang Sukowati : Media Penelitian Dan Pengembangan*, *6*(1), 98–108. https://doi.org/10.32630/sukowati.v6i1.329

Asyfani, Y., Manfaati Nur, I., Fathoni Amri, I., Yunanita, N., Hikmah Nur Rohim, F., Aura Hisani, Z., & Anggun Lestari, F. (2024). Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Kepadatan Penduduk Menggunakan Metode Hierarchical Clustering Info Artikel. *Journal of Data Insights*, *2*(1), 1–8. Retrieved from http://103.97.100.158/index.php/jodi/article/view/158

Boyko, N. I., & Tkachyk, O. A. (2023). Hierarchical clustering algorithm for dendrogram construction and cluster counting. *Informatics and Mathematical Methods in Simulation*, *13*(1–2), 5–15. https://doi.org/10.15276/imms.v13.no1-2.5

Dalimunthe, M. (2021). Mengatasi Kemiskinan dalam Islam (Kajian Al-Quran dan Al-Hadis). *Journal of Islamic Law El Madani*, *1*(1). http://journal.marwah-madani-riau.id/index.php/JILE/article/view/8

Dinata, E., & Syaputra, H. (2020, September 23). *Penerapan Metode Agglomerativ Hirarchical Clusturing untuk Klasifikasi Dokumen Skripsi*. https://www.semanticscholar.org/paper/PENERAPAN-METODE-AGGLOMERATIV-HIRARCHICAL-UNTUK-Dinata-Syaputra/ed25f7bb2e17200bc58c5d8089585519b3f9609c

Fuady, M., Fuady, M. R. F., & Aulia, F. (2022). Kemiskinan Multi Dimensi dan Indeks Pembangunan Manusia di Indonesia. *Tataloka*, *24*(4), 330–337. https://doi.org/10.14710/tataloka.24.4.330-337

Hasan, Y. (2024). Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means Dan DBSCAN. *JITET(Jurnal Informatika Dan Teknik Elektro Terapan)*, *06*(01), 60–74. Retrieved from https://ejournal.ust.ac.id/index.php/KAKIFIKOM/article/view/3938

Hilmawan, R., Aprianti, Y., Vo, D. T. H., Yudaruddin, R., Bintoro, R. F. A., Fitrianto, Y., & Wahyuningsih, N. (2023). Rural development from village funds, village-owned enterprises, and village original income. *Journal of Open Innovation: Technology, Market, and Complexity*, *9*(4), 100159. https://doi.org/10.1016/j.joitmc.2023.100159

Karmanita, D., Defit, S., & Sumijan. (2024). Metode K-Means dalam Mengukur Tingkat Pemahaman Materi Mata Kuliah dan Penilaian Mahasiswa Prodi Informatika. *Jurnal KomtekInfo*, *101*(3), 132–138. https://doi.org/10.35134/komtekinfo.v11i3.551

Larasati, N. F., Fauziyyah, N., & Anugrah, R. (2024). Analisis Ketimpangan Kemiskinan di Provinsi Jawa Tengah dengan Metode Agglomerative Hierarchical Clustering. *Seminar Nasional Pendidikan Matematika Dan Matematika*, *10*. https://doi.org/10.21831/pspmm.v10i2.342

Muflihan, Y., Retnawati, H., & Kristian, A. (2022). Analisis cluster dengan metode hierarki untuk pengelompokan sekolah menengah atas berdasarkan rapor mutu sekolah di Kabupaten Nagan Raya. *Measurement In Educational Research (Meter)*, *2*(1), 22–33. https://doi.org/http://dx.doi.org/10.33292/meter.v2i1.155

Nikum, B., & Yuliansyah, H. (2024). Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation. *ILKOM Jurnal Ilmiah*, *16*(3), 330–342.

Olexandr TKACHYK, L. P. N. U. (2023). Applying Data Clustering Methods for Creating Targeting User Groups for Real Estate. *Herald of Khmelnytskyi National University. Technical Sciences*, *319*(2), 300–207. https://doi.org/10.31891/2307-5732-2023-319-1-300-307

Pradana, M. G., & Ha, H. T. (2021). Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering. *Journal of Applied Data Sciences, 2*(1), 19–25. https://doi.org/10.47738/jads.v2i1.18

Provinsi Jawa Tengah, B. P. S. (2024). *Provinsi Jawa Tengah Dalam Angka 2024* (Tahunan). https://jateng.bps.go.id/id/publication/2024/02/28/980d120f5be18d6400c48b16/provinsi-jawa-tengah-dalam-angka-2024.html

Rahmawati, N., Prasetyanto, P. K., & Islami, F. S. (2022). Pengaruh Indeks Pembangunan Manusia (Ipm), Upah Minimum Regional (UMR), Dan Tingkat Pengangguran Terbuka Terhadap Jumlah Penduduk Miskin Di Indonesia Tahun 2017-2021 (Studi Kasus 34 Provinsi Di Indonesia). *Convergence: The Journal of Economic Development*, *4*(1), 19–31. https://doi.org/10.33369/convergencejep.v4i1.23358

Repiská, R., Grisáková, N., & Štetka, P. (2022). Hierarchical Clustering Based On International Sustainability Indices Of Eu Countries 1 Introduction. *EKONOMICKÉ ROZHĽADY – ECONOMIC REVIEW*. *51*(2), 149–

170. http://dx.doi.org/10.53465/ER.2644-7185.2022.2.149-170

Sharma, M. M. (2022). Comparison Of Clustering Algorithms : A Practical Approach. *OORJA, 20*(1), 64–74.

Thakare, V. S., & Sonawane, J. S. (2023). An Effective Unsupervised Machine Learning Technique and Research Challenges. *International Journal of Science and Research (IJSR)*, *12*(5), 2141–2143. https://doi.org/10.21275/SR23523214829

Tripathi, S., Muhr, D., Brunner, M., Jodlbauer, H., Dehmer, M., & Emmert-Streib, F. (2021). Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing. *Frontiers in Artificial Intelligence*, *4*(June), 1–20. https://doi.org/10.3389/frai.2021.576892

Wardani, A. N. K., Pasya, D. J., & Juliono, J. (2023). Pengaruh Indeks Pembangunan Manusia (IPM) Terhadap Tingkat Kemiskinan di Provinsi Lampung. *Jurnal Ilmiah Ekonomi Manajemen Jurnal Ilmiah Multi Science*, *14*(01), 06–11. https://doi.org/10.52657/jiem.v14i01.2045

Zulkifilu, M., & Yasir, A. (2022). About Some Data Precaution Techniques for K-Means Clustering Algorithm. *UMYU Scientifica*. *1*(1), 12–19. https://doi.org/10.56919/usci.1122.003