SENTRI: Jurnal Riset Ilmiah
https://ejournal.nusantaraglobal.ac.id/index.php/sentri
Email: sentri@nusantaraglobal.ac.id
DOI: https://doi.org/10.55681/sentri.v5i8.7040

Vol. 5, No. 8, August 2026
e-ISSN: 2963-1130
pp. 4243-4253

Analysis of District Clustering in East Priangan, West Java, Based on Health Education and Infrastructure Using the K-Means Method

Yuga Mudzadid Fiqri[1*], Atika Nurani Ambrawati[1]
[1]Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Indonesia

*Corresponding Author's e-mail: sizadid20@gmail.com

Article History:
Received: July 30, 2026
Revised: August 18, 2026
Accepted: August 31, 2026

Keywords:
K-Means Clustering; Sub-district Clustering; East Priangan

Abstract: *Development disparities between regions remain an issue that affects the equitable distribution of educational, health and infrastructure facilities. Differences in the availability of these facilities lead to variations in service levels and development across sub-districts. This study aims to cluster sub-districts in the East Priangan region of West Java based on educational, health and infrastructure characteristics using the K-Means Clustering method. The data used are from 2024, obtained from the Central Statistics Agency (BPS) of districts/cities in the East Priangan region, comprising 13 research variables and 132 sub-districts as the objects of analysis. Prior to the clustering process, the data was tested using the Kaiser-Meyer-Olkin (KMO) and multicollinearity tests. The test results showed a KMO value of 0.846, indicating that the data was suitable for further analysis and that no multicollinearity issues were found. Determining the optimal number of clusters using the Elbow method yielded two clusters as the optimal number. The clustering results showed that Cluster 1 comprised 30 sub-districts with relatively higher facility availability, whilst Cluster 2 comprised 102 sub-districts with relatively lower facility availability. The distance between cluster centres of 202.258 indicated a fairly clear difference in characteristics between the two clusters. It is hoped that the results of this study can serve as a basis for the formulation of more effective and targeted development policies.*

How to cite: Fiqri, Y. M., & Ambrawati, A. N. (2026). Analysis of District Clustering in East Priangan, West Java, Based on Health Education and Infrastructure Using the K-Means Method. *SENTRI: Jurnal Riset Ilmiah*, *5*(8), 4243–4253. https://doi.org/10.55681/sentri.v5i8.7040

INTRODUCTION

Development is a key indicator of a region's progress, but disparities among regions remain a problem due to a lack of equitable development. [1] Development, at its core, is a process of continuous change toward a better future, based on certain norms. The primary goal of development is to improve the overall standard of living of society, encompassing aspects such as education, health, economic well-being, and social justice. [2] In addition, unequal access to education, health care, and infrastructure is also a problem in every region. [3] Differences in the characteristics of regencies and cities in terms of economic structure, local tax base, and level of urbanization lead to significant variations in their capacity to finance development.[4] West Java, as the most populous province in Indonesia, has a highly complex fiscal landscape East Priangan is one of the main regions in West Java Province, consisting of several regencies and cities, namely Garut Regency, Ciamis Regency, Tasikmalaya Regency, Pangandaran Regency, Tasikmalaya City, and Banjar City.[5]

Although the East Priangan region has considerable development potential, disparities in the availability of educational and health facilities and infrastructure remain evident among the subdistricts in the region. These differences have not yet been systematically classified, making it difficult for local governments to formulate development policies that reflect the specific characteristics of each area. As a result, development programs are often implemented using a one-size-fits-all approach, which can reduce their effectiveness and contribute to inter-regional disparities.

Therefore, an analysis is needed to identify similarities among the subdistricts in East Priangan. Clustering is an appropriate approach because it groups areas with similar characteristics, simplifies multidimensional data, and facilitates interpretation. Among the various clustering methods, K-Means was chosen for its ability to classify quantitative data into homogeneous groups based on similarities while producing clear distinctions between clusters. This method is suitable for analyzing indicators of education, health, and infrastructure and provides results that are easy to interpret for regional planning.

This study aims to classify 132 subdistricts in the East Priangan region of West Java based on indicators of education, health, and infrastructure using the K-Means clustering method. The resulting clusters are expected to provide an overview of regional development characteristics and to support local governments in formulating development policies that are more targeted, effective, and equitable.

LITERATURE REVIEW

Elbow Method

The Elbow Method is a technique used to determine the optimal number of clusters by observing the pattern of decline in a graph, where a point resembling an "elbow" is formed. The comparison value is obtained by calculating the Sum of Squared Errors (SSE) for each number of clusters. Therefore, the optimal number of clusters is determined at the point where the decrease in SSE begins to level off (forming an "elbow"), based on the SSE equation used [1].

Cluster Analysis

Cluster analysis is an effective statistical technique for grouping regions based on the similarity of specific characteristics. The K-Means clustering method is widely used in socio-economic and fiscal research because of its ability to form homogeneous groups by minimizing the distance between observations. This technique enables researchers to identify hidden patterns within the data and produce a more objective regional classification than conventional descriptive approaches [2].

Euclidean Distance

Euclidean Distance is a method for calculating the distance between two points in Euclidean space using Cartesian coordinates. It is commonly employed in clustering algorithms to measure the similarity or dissimilarity between observations. The shorter the Euclidean distance, the greater the similarity between the data points [3].

K-Means Clustering

K-Means Clustering is a non-hierarchical clustering method that partitions objects into one or more groups based on their characteristics. Objects with similar characteristics are assigned to the same cluster, while those with different characteristics are assigned to different clusters. The K-Means algorithm aims to group existing data into several clusters such that objects within the same cluster are more similar to one another than to those in other clusters [4].

5. Variables

The data used in this study consist of district- and municipality-level data obtained from the Central Bureau of Statistics (Badan Pusat Statistik/BPS) in each district and municipality. The dataset includes research indicators related to education, health, and infrastructure.

The education sector consists of the following variables: Kindergarten/Islamic Kindergarten (TK/RA/BA), Elementary School/Islamic Elementary School (SD/MI), Junior High School/Islamic Junior High School (SMP/MTs), Senior High School/Islamic Senior High School (SMA/MA), and Vocational High School (SMK). The health sector includes only one variable, namely the number of midwives. The infrastructure sector consists of the number of hospitals, public health centers (Puskesmas), mosques, prayer rooms (Mushola), Protestant churches, Catholic churches, and hotels. The variables used in the analysis represent the total number of each facility or institution within each district and municipality.

METHOD

The data used in this study consists of data from each city and regency in the Prianan region of East West Java, presented as numerical values for each variable; this data was officially obtained from the Central Statistics Agency (BPS) in each regency and city in 2024.

The variables used are numerical variables covering the aspects of education, health, and infrastructure.

RESEARCH DATA AND VARIABLES

This study covers indicators representing data from various sectors, such as health, education, and infrastructure, at the subdistrict level in the regencies and cities of the East Priangan region, West Java. These variables were selected because they are considered capable of reflecting the level of regional development and serve as the basis for the clustering process using the K-Means method. The variables used in this study are as follows

Table 1. Research Variables

Variables	Description
X1	Amount TK/RA/BA
X2	Amount SD/MI
X3	Amount SMP/MTS
X4	Amount SMA/MA
X5	Amount SMK
X6	Amount Midwife
X7	Amount Hospital
X8	Amount Health Center
X9	Amount Mosque
X10	Amount Prayer Room

X11	Amount Protestant Church
X12	Amount Catholic Church
X13	Amount Hotel

Research Stages

1. Collection of data obtained from the respective district and city BPS offices in each region.
2. Establishing cluster assumptions
3. Determining the number of clusters using the elbow method based on the potential indicators of villages in East Priangan, West Java.
4. Determining the cluster centres
5. Determining the distances of objects from the center point
6. Grouping based on minimum distance
7. Interpretation of clustering results
8. Drawing conclusions and recommendations

Sample Adequacy Test

Data analysis using factor analysis is a method used to identify the dominant factors in explaining a specific problem. The KMO (Kaiser-Meyer-Olkin) test is conducted to determine the suitability of the factor analysis to be performed. The KMO value and Bartlett's test statistic must be greater than 0.5, with a significance level of less than 0.05. Factors can be analyzed in the next stage if these criteria are met.[6]

Algorithm K-Means

K-Means is a method for clustering data based on distance, using only numerical characteristics.[7] K-Means clustering is a data clustering method that does not use a hierarchical structure. This method aims to group data into several clusters so that data with similar characteristics are placed in the same cluster. This method seeks to group existing objects into one or more clusters based on their characteristics. As a result, objects with the same characteristics will be grouped into the same cluster, while objects with different characteristics will be grouped into different clusters.[8]

1. Determine the value of k and the center of mass for each cluster
2. Calculate the Euclidean distance between each object and the center of mass. The following equation can be used to calculate the Euclidean distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

Notes:
d(x, y) = Euclidean distance
i = each data point 1, 2, ..., n
n = the number of data points
x = cluster centre data
y = data for the attribute of the cluster centre
xi = data for the i-th cluster centre

yi = data for the i-th data point

3. Clustering Based on Minimum Distance

Clustering based on minimum distance is a process for determining the membership of each data object in a cluster by comparing the object's distance to all existing cluster centers, and then assigning it to the cluster with the shortest distance

.

$$d(x,c) = \sqrt{\sum_{j=1}^{p} \left(x_j - c_j\right)^2} \quad (2)$$

xj = the value of the jth variable in the object
cj = the mean value of the jth variable
p = the number of variables

Determining the Number of Clusters

The Elbow Method is a technique used to determine the optimal number of clusters by observing the pattern of decline in a graph, where a point resembling a right angle is formed. Comparison values are obtained by calculating the Sum of Squared Errors (SSE) for each number of clusters. Therefore, the optimal number of clusters is determined at the point where the decline in SSE begins to level off (forming a "right angle"), in accordance with the SSE equation used.[9]

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in X_k} |x_i - c_k|^2 \quad (3)$$

Notes:
K = the cth cluster
Xi = the distance of the ith data point
ck = the centre of the ith cluster

Table 2.Descriptive Statistics

Descriptive Statistics					
Variable	N	Minimum	Maksimum	Mean	Std.Deviation
TK/RA/BA	132	4.00	91.00	22.7955	15.51593
SD/MI	132	5.00	77.00	26.7273	16.52160
SMP/MTS	132	2.00	37.00	12.9318	6.97673
SMA/MA	132	.00	15.00	4.8636	3.200251
SMK	132	00	15.00	3.5606	2.38717
Midwife	132	5.00	233.00	31.439	31.34863
Hospital	132	.00	2.00	.2197	.52878
Health Center	132	.00	3.00	1.0682	.79303
Mosque	132	5.00	451.00	98.5227	85.26863
Player Room	132	2.00	522.00	69.8561	85.67025
Protestant Church	132	.00	9.00	.2803	1.12782
Catholic Church	132	.00	3.00	.0909	.37993
Hotel	132	.00	54.00	1.4167	5.48842

Based on Table 2, it can be seen that there are differences in the number of educational, health, and infrastructure facilities across the 132 subdistricts in the East Priangan region. Facilities such as mosques, prayer rooms, elementary schools (SD/MI), and midwives' clinics have relatively high average values, while hospitals, churches, and hotels have low averages. Furthermore, the significant difference between the minimum and maximum values indicates disparities in the availability of facilities across subdistricts.

Cluster Assumption Test

Table 3. KMO Value

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,846
Bartlett's Test of Sphericity	Approx. Chi-Square	1002,759
	df	78
	Sig.	,000

The Kaiser-Meyer-Olkin (KMO) value of 0.846 indicates that the data have a very good level of sample adequacy for factor analysis or further analysis. This value falls within the range of 0.80–0.89, meaning that the relationships among the variables are strong enough to make the data suitable for analysis.

Determining the Optimal Number of Clusters Using the Elbow Method

Table 4. SSE Value of Each Cluster

Amount Cluster (k)	**SSE**
1	2121465,7
2	1165277,2
3	852962,9
4	655544,3
5	488634,7
6	399811,5
7	336754,6
8	281124,1
9	234822,8
10	212863,4

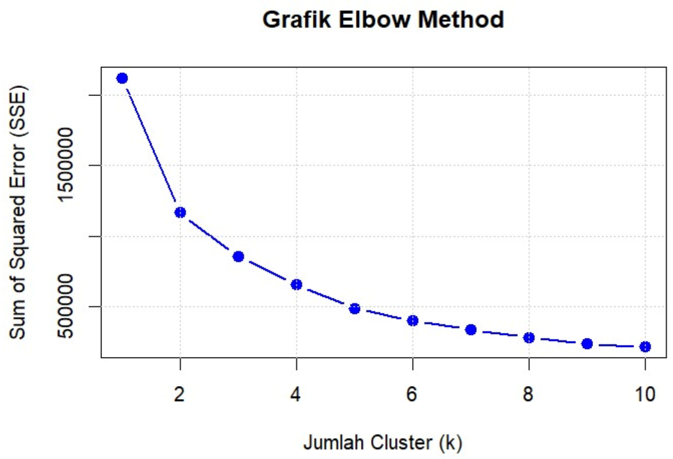

Figure 1. Graph of the Elbow Method for Determining the Optimal Number of Clusters

Based on the Elbow Method graph, the SSE value decreases very sharply from k = 1 to k = 2. After k = 2, the decrease in SSE tends to be more gradual, which means that increasing the number of clusters does not result in a significant decrease in SSE. Therefore, the optimal number of clusters selected is 2.

Clustering Using The K-Means Method

Once the optimal number of clusters has been determined, the next step is to find the cluster center values for each cluster. The cluster center values are obtained by averaging the values within each cluster, starting with a random selection for the initial cluster center. The clustering process ends when the new cluster center value is equal to the previous one. The following are the final cluster center values obtained.

Table 5. Final Cluster Centers

Variable	Cluster	
	1	2
TK/RA/BA	.94976	-.52516
SD/MI	1.00297	-.55458
SMP/MTS	.98566	-.54501
SMA/MA	.90626	-.50111
SMK	.72775	-.40240
Midwife	.77329	-.42758
Hospital	.30879	-.17074
Health Center	.58476	-.32334
Mosque	.93083	-.51470
Player Room	.73656	-.40728
Protestant Church	.29856	-.16508
Catholic Church	.20873	-.11542
Hotel	.13342	-.07377

Cluster 1	:	Cluster 1 consists of city centers or subdistricts with high population density, characterized by high indicator scores across the board. In terms of education indicators, this cluster clearly excels, with an abundance of school facilities at every level, particularly at the elementary school level. Its health indicators are also very strong, as evidenced by the high availability of medical personnel. These strengths are reinforced by a very extensive infrastructure sector, ranging from well-established physical health facilities and economic support facilities to a very dense network of places of worship.
Cluster 2	:	Cluster 2 consists of rural areas or buffer zones characterized by low and limited levels of development across all sectors. In terms of education indicators, access to formal education tends to be very limited at all levels, particularly at the high school level, where such services remain scarce. The health sector also reflects limitations in primary care services due to the relatively small number of available medical personnel. These limitations are most evident in the public infrastructure sector, where provision remains uneven; both physical health facilities and the number of places of worship are still very limited and require development.

Cluster Membership

Table 6. Eastern Priangan District West Java

Cluster 1	**Cluster 2**		
Banjar	Cimerak	Sukarame	Sukaresmi
Purwaharja	Cijulang	Cigalontang	Samarang
Cipatujah	Cigugur	Leuwisari	Wanaraja
Cikalong	Langkap lancar	Padakembang	Sucinaraja
Pancatengah	Parigi	Sariwangi	Pangatikan
Cikatomas	Sidamulih	Sukaratu	Karangtengah
Sodonghilir	Pangandaran	Cisayong	Leles
Kawalu	Kalipucang	Sukahening	Lewigoong
Tamansari	Padaherang	Rajapolah	Kersamanah
Cibeureum	Mangunjaya	Jamanis	Cibiuk
Mangkubumi	Banjarsari	Ciawi	Bluburlimban
Pakanjeng	Banjaranyar	Kadipaten	Salawi
Cikelet	Lakbok	Pagerageung	Culamega
Cisompet	Prwadadi	Sukaresik	Bojonggambir
Singajaya	Pamarican	Purbaratu	Taraju
Cikajang	Cidolog	Tawang	Salawu
Banjarwangi	Cimaragas	Cihideung	Puspahiang
Cilawu	Cijeunjing	Indihiang	Tanjungjaya
Bayongbong	Cisaga	Bangunsari	Sukaraja
Cisurupan	Tambaksari	Cipedes	Salapo
Pasirwangi	Rancah	Cisewu	Jatiwaras

Tarogong kidul	Rajadesa	Caringin	Parungponteng
Tatogong kaler	Sukadana	Talegong	Bantarkalong
Garut Kota	Ciamis	Bungbulang	Bojongasih
Karangpawitan	Baregbeg	Mekarmukti	Cineam
Sukawening	Cikoneng	Pamulihan	Karangjaya
Banyuresmi	Sindangkasih	Pameungpeuk	Manonjaya
Cibatu	Cihaurbeuti	Cibalong	Gunungtanjung
Kadungora	Sadananya	Peundeuy	Singaparna
Malangbong	Cipaku	Cihurip	Mangunreja
	Jatinagara	Cigedug	Panawangan
	Karangnunggal	Panumbangan	Kawali
	Cibalong	Pataruman	Lumbung
	Sukamantri	Lengensari	Panjalu

Results of the grouping of 132 subdistricts in the East Priangan region into two clusters based on the availability of educational, health, and infrastructure facilities. Cluster 1 has a higher average score on nearly all variables, indicating that the subdistricts in this cluster have a relatively more comprehensive range of facilities. Meanwhile, Cluster 2 has lower average scores on most variables, indicating that subdistricts in this cluster have a relatively lower level of facility availability. These clustering results highlight differences in facility characteristics among subdistricts in the East Priangan region and can serve as a basis for establishing priorities for more equitable development.

Table 7. Initial Cluster Centers

Variabel	Cluster	
	1	2
TK/RA/BA	2.14003	-1.01802
SD/MI	.86388	-1.13350
SMP/MTS	1.01311	-.99356
SMA/MA	1.29160	-.58193
SMK	1.02188	-.23484
Midwife	1.48525	-.74770
Hospital	1.47566	1.47566
Health Center	2.43598	-.08598
Mosque	4.13373	-1.06162
Player Room	5.27772	-.72203
Protestant Church	3.29813	1.52480
Catholic Church	5.02489	-.239228
Hotel	10628	9.58077

Based on Table 7, the Initial Cluster Centers values indicate the initial cluster centers used as the starting point for the K-Means process. The values for Cluster 1 are generally higher than those for Cluster 2, indicating that the initial center of Cluster 1 represents subdistricts with greater availability of facilities, while Cluster 2 represents subdistricts with fewer facilities. These values are still preliminary and will be updated iteratively until stable final cluster centers are obtained.

Distance Between Clusters

Table 8. Distance Between Clusters

Distances between Final Cluster Centers		
Cluster	1	2
1		202.258
2	202.258	

Based on the analysis results, the distance between the centers of the final clusters was found to be 202.258. This value indicates the degree of difference in characteristics between Cluster 1 and Cluster 2. The greater the distance between clusters, the more distinct the characteristics of the members within each cluster.

A distance value of 202.258 indicates that the two clusters formed have quite striking differences, meaning that the clustering process successfully separated the subdistricts in the East Priangan region into two groups with distinct characteristics based on the variables used in this study.

CONCLUSION

Based on the research findings, the K-Means clustering method successfully grouped 132 subdistricts in the East Priangan region of West Java into two clusters based on indicators of education, health, and infrastructure. The clustering results indicate differences in characteristics among subdistricts based on the level of facility availability.

Cluster 1 consists of 30 subdistricts with higher average scores on nearly all indicators, reflecting subdistricts with relatively more comprehensive availability of education, health, and infrastructure facilities. Meanwhile, Cluster 2 consists of 102 subdistricts with lower average scores on most indicators, indicating subdistricts with relatively more limited availability of facilities. These clustering results reveal variations in the level of facility availability among subdistricts in the East Priangan region and are expected to serve as a basis for local governments in formulating more targeted development policies and supporting the equitable distribution of facilities throughout the region**.**

REFERENCES

[1] G. Rohma, U. Asyafiiyah, A. Widyastuti, and F. Andriani, “Analisis Wilayah Prioritas Pembangunan di Provinsi Jawa Timur Berdasarkan Indikator Sosial Menggunakan Metode K-Means Clustering,” vol. 5, no. 1, pp. 32–48, 2025.

[2] M. Amelia, A. Faqih, and A. R. Rinaldi, “Penerapan Metode K-Means Clustering dalam Pemetaan Kemiskinan Kabupaten/Kota di Tepat,” *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, vol. 13, no. 2, 2025. doi: http://dx.doi.org/10.23960/jitet.v13i2.6231.

[3] V. Azizah and D. Widiyanto, “Analisis Hubungan Tingkat Perkembangan Ketimpangan Wilayah di Indonesia,” vol. 20, 2025. doi: 10.20961/region.v20i1.86894.

[4] B. Besaran, A. Dan, and P. Per, “Analisis K-Means Cluster untuk Pengelompokan Wilayah Jawa Barat Berdasarkan Besaran APBD dan Pendapatan Per Kapita,” *Global Research Innovation Journal*, vol. 1, no. 3, pp. 2925–2938, 2025. [Online]. Available: https://journaledutech.com/index.php/great

[5] P. Seminar and N. Kemahasiswaan, “Analisis Komparatif Sektor Unggulan di Priangan Timur,” vol. 2, pp. 159–174, 2025.

[6] D. A. I. Cahya, D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” vol. 9, no. 3, 2019.
[7] “Klasifikasi Status Gizi Balita untuk Deteksi Dini Stunting Menggunakan Metode K-Nearest Neighbor,” *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 4, pp. 2777–2786, 2024.
[8] A. Zaki and A. Sembe, “Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM),” *Journal of Mathematics, Computation, and Statistics*, vol. 5, no. 2, pp. 163–176, 2022. [Online]. Available: http://www.ojs.unm.ac.id/jmathcos
[9] T. Bah et al., “Analisis Faktor-Faktor yang Mempengaruhi Revisit Intention (Minat Kunjung) Wisatawan pada Objek Air,” vol. 5, no. 1, pp. 25–34, 2023.
[10] A. N. Oktaviani and A. N. Ambarwati, “Implementation of K-Means Cluster for Districts or Cities in West Java Province Based on Unemployment Indicators,” vol. 18, no. 1, pp. 877–886, 2025.