EKSAKTA
Journal of Sciences and Data Analysis

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art11

Research Article

Application of the K-Means ++ Method for Grouping Health Services Based on Districts in West Java Province

Indah Manfaati Nur [2]*, Abdurakhman[1]

[1] Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada
[2] Statistics, Faculty of Sciences and Agricultural Technology, Universitas Muhammadiyah Semarang

[*] *Corresponding author:indahmanfaatinur@mail.ugm.ac.id*

Received: 18 December 2023; Accepted:25 April 2024; Published: (96-102)

Abstract: A healthy and prosperous life is one of the goal points listed in the Sustainable Development Goals (SDGs). To create a healthy life, support is needed in the form of equal health facilities and services for all people in all provinces of Indonesia. In fact, there are still provinces that have low levels of health service facilities. West Java is ranked last with low health service conditions. Grouping efforts are needed to identify cities or districts in West Java that deserve priority for handling health facilities. In this study, the K-Means++ method was used to group health facilities based on cities or districts in West Java Province. Based on the grouping results, 3 clusters were obtained, namely cluster 1 for groups with low health facilities with 18 cities/regencies as members, cluster 2 for groups with moderate health facilities with 7 cities/regencies, and cluster 3 for groups with high health facilities with 2 cities/regencies.

Keywords: K-Means++, Clustering, Health Facilities, Jawa Barat

Introduction

Sustainable Development Goals (SDGs) are a series of sustainable development goals and initiatives designed by member countries of the United Nations (UN) in 2015. This program is expected to achieve various sustainable development targets until 2030. SDGs contain 17 sustainable development goals, one of the goal points is to create a healthy life and improve the welfare of the entire population which is written in the third point. In order to create a healthy life, there needs to be support and improvement in health services.

Health services are an activity organized by the government to provide care to individuals and communities, with a focus on preventing and treating disease at both the individual and community levels [1]. Health services have an important role in ensuring the welfare of society as a whole, thereby contributing to the productivity and sustainability of society. Ease of access to health services provides opportunities for early detection and treatment of disease, reducing the impact of disease, and improving overall quality of life. Businesses in providing health services also have a role in creating an environment that supports healthy lifestyles, providing health education, and encouraging medical research. Therefore, attention to health services is an important step in strengthening and improving the quality of society.

Based on the ranking of health service conditions from 34 provinces, West Java is in last place with 15.43 points [2]. The lack of health services is a serious problem that can have an impact on people's welfare [3]. Several factors that can cause low levels of health services in this region include the unequal distribution of health facilities and medical personnel, especially between urban and rural areas [4][5][6]. To improve the quality of health services in West Java, a regional grouping of these indicators is needed. This is needed to help increase the accessibility of health services, especially in less developed areas. Grouping can be done using clustering methods, one of which is the K-Means++ method.

K-Means++ is a development of the K-Means method, which uses an approach to overcome cases of uncertainty that can be generated by the standard K-Means algorithm. The difference between the K-Means++ and K-Means methods lies in the way the data is entered. In the K-Means method algorithm, the first data entered into an empty cluster is done randomly. Meanwhile, in the K-Means++ algorithm, entering data into each cluster uses a certain distance calculation, so it is no longer a random system. Much research has been conducted on K-Means++ previously. Research by Nugroho and Adhinata (2022) using the K-Means++ method for Clustering Covid-19 Data on the Island of Java obtained several clustering models

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art11

using different k, the optimum k value was k=5 which resulted in a silhouette coefficient value of 0.882. In this research, the results obtained from K-Means++ were superior in providing information on the extent of the spread of the Covid-19 virus compared to using the regular K-Means method.

Based on this background, this article will discuss the application of K-Means++ in grouping health service data in West Java Province. The aim of this research is to obtain clusters based on health services, so that it can help certain parties to equalize health services in West Java Province.

Materials and Methods

Materials

The data used is secondary data obtained from the official website https://opendata.jabarprov.go.id/id. This research uses data regarding health facilities in 26 districts or cities in West Java Province with 8 indicators which can be seen in Table 1.

Table 1. Health Facility Indicators

Indicator	Description
General Hospital	Total number of public hospitals in districts/cities
Special Hospital	Total number of special hospitals in districts/cities
Public Health Center	Total number of Community Health Centers (Puskesmas) in districts/cities
Integrated Service Post	Total number of Integrated Service Post (Posyandu) in districts/cities
Midwife	Total number of midwives in the district/city
Nurse	Total number of nurses in the district/city
Medical Specialist	Total number of specialist doctors in the district/city
General Practitioners	Total number of general practitioners in the district/city

K-Means++

K-Means++ is an algorithm used to determine an initial value or "seed" as an approach to overcome cases of uncertainty that can be generated by the standard K-Means algorithm [7]. The goal of K-Means++ is to spread the initial centric points as far apart as possible from each other, thus making the solution easier. The K-Means++ method has several advantages, including being able to produce solutions that are superior to the standard K-Means solution and providing better initialization for the K-Means algorithm so that it can improve the quality of data grouping. Apart from that, the K-Means++ method can also reduce the convergence time to the optimal solution [8].

The steps in the K-Means++ algorithm are as follows [1][2]:

1. Determine the number of k values, namely the number of clusters that will be used. In this study, the k value of 3 clusters was used.
2. Determine the cluster center point (centroid) using the following equation:

$$K = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (2.1)$$

Information:
$D(x)^2$ = Euclidean distance
$\sum_{x \in X} D(x)^2$ = Sum of Euclidean Distances

3. Calculate the Euclidean distance of each data object using the following equation:

$$d_{(x,y)} = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2} \quad (2.2)$$

Information:
$d_{(x,y)}$ = Distance of data from x to cluster center y
x_i = Data x on the i[th] observation
y_i = The y-center point of the i[th] observation
n = Lots of observations

$$V_{ij} = \frac{1}{N_i \sum_{k=0}^{N_i} X_{kjZ}} \quad (2.3)$$

Information:
V_{ij} = Average centroid of the I[th] cluster for the j[th] variable
N_i = Number of members of the i[th] cluster
i, k = Index of clusters

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art11

j = Variable index
X_{kj} = The k[th] data value of the j[th] variable for the cluster

4. Grouping based on the minimum distance to the centroid.
5. Update the centroid point by finding the average of each cluster.
6. Repeat steps 2 and 3 until all objects have not moved.

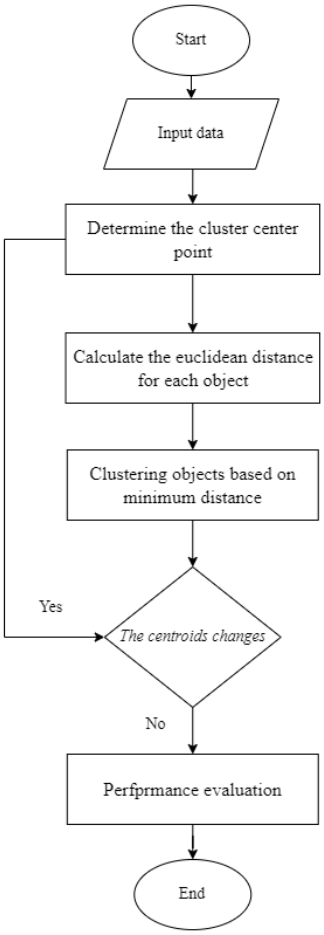

Figure 1. Flowchart K-Means++

Sum Of Squared Error

The Sum Of Squared Error (SSE) value is a statistical measure that shows how far the dependent variable deviation point is from the predicted value [11]. SSE is used as a measure of model accuracy and is calculated by summing the squared differences between the observed values and the predicted values of the dependent variable. The lower the SSE value, the better the model created, while the higher the SSE value, the worse the model created. The SSE value can be found using the following equation formula:

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2.4)$$

Keterangan:
SSE = Sum Of Squared Error (SSE)
y_i = Observation value
$\hat{y}$ = Predicted value

Result and Discussion

Descriptive Analysis

Descriptive analysis is a statistical method used to provide a summary and clear picture of the basic properties of a set of data. Table 2 shows the Mean, Standard Deviation, Minimum and Maximum values for each indicator.

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art11

Table 2. Descriptive Analysis

Indikator	Mean	Standar Deviasi	Max	Min
General Hospital	12	11	48	1
Special Hospital	2	3	15	0
Public Health Center	41	21	101	10
Integrated Service Post	1926	1267	5133	200
Midwife	1021	548	2670	198
Nurse	2211	1615	8110	494
Medical Specialist	317	345	1277	21
General Practitioners	235	230	905	11

Determination of Multiple Clusters

Determining the number of clusters needs to be done before running the clustering process. The number of clusters can be determined randomly or using certain methods, such as the elbow method. The elbow method uses the total wss (within sum square) value as a criterion for determining the optimal cluster.

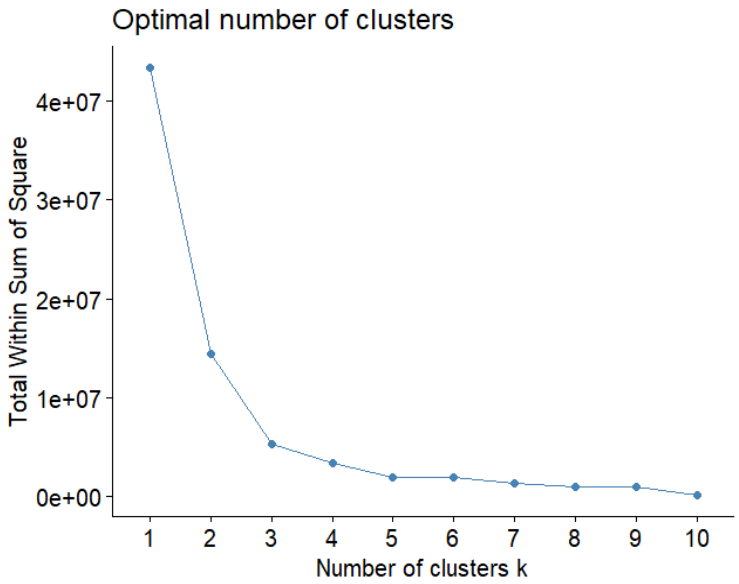

Figure 2 Elbow Method Results

Figure 2 shows a line that experiences a fault forming an elbow when k=3. Based on these results, 3 clusters will be tested with cluster 1 for the group with a low level of health facility service, cluster 2 for the group with a medium level of health facility service, and cluster 3 for the group with a high level of health facility service.

Metode K-Means++ Clustering

The following are the results of grouping City/Regency into three clusters based on health services in West Java province using the K-Means++ method.

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art11

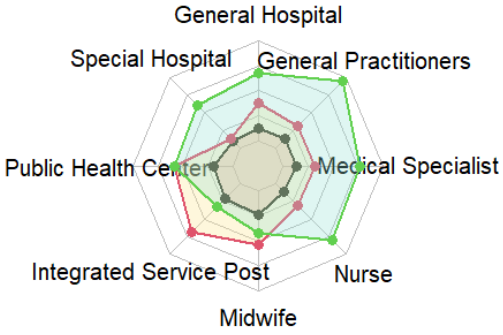

Figure 3. Indicator Characteristics of Each Cluster

Based on Figure 3, it can be interpreted that each cluster has its own characteristics. Cluster 3 has the highest number of Community Health Centers, Special Hospitals, General Hospitals, General Practitioners, Specialist Doctors and Nurses compared to other clusters. Cluster 2 has the highest number of Integrated Service Post and Midwives compared to other clusters. Apart from that, all indicators in cluster 1 have low values compared to other clusters.

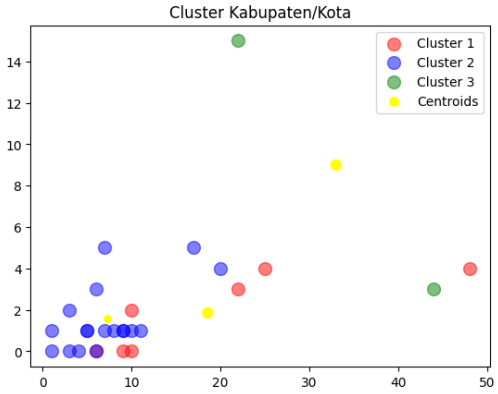

Figure 4. Clustering Results

Figure 4 shows the results of grouping City/Regency in West Java Province based on health facilities. The grouping results obtained were 3 clusters with their respective centroids. Figure 5 is the result of cluster visualization for each City/Regency with light blue indicating cluster 1, blue indicating cluster 2, and dark blue indicating cluster 3.

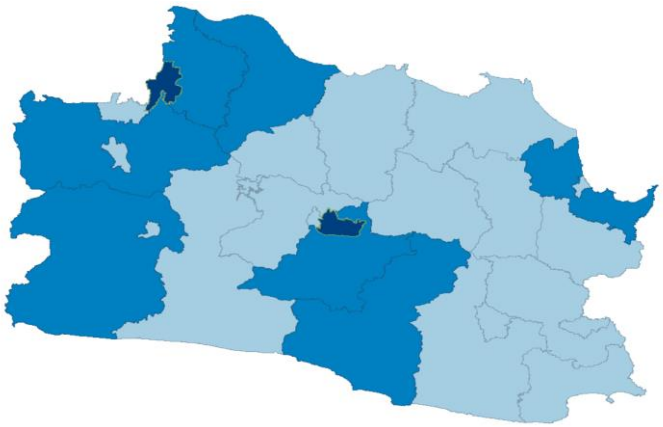

Figure 5. Clustering Map of Health Facilities by City/Regency

From the results of the K-Means++ cluster, the number of regencies/cities in the low cluster is 18 City/Regency, the medium cluster is 7 City/Regency, and the high cluster is 2 City/Regency. This can be

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art11

interpreted that by using the K-Means++ method the grouping results are dominated by low clusters. Table 3 below shows the details of the City/Regency in each cluster.

Table 3. City/Regency Cluster Results

Cluster	Number of Members	City/Regency
1	18	Cianjur Regency, Tasikmalaya Regency, Ciamis Regency, Kuningan Regency, Majalengka Regency, Sumedang Regency, Indramayu Regency, Subang Regency, Purwakarta Regency, Bandung Barat Regency, Pangandaran Regency, Bogor City, Sukabumi City, Cirebon City, Depok City, Cimahi City, Tasikmalaya City, Banjar City
2	7	Bogor Regency, Sukabumi Regency, Bandung Regency, Garut Regency, Cirebon Regency, Karawang Regency, Bekasi Regency
3	2	Bandung City, Bekasi City

Model Evaluation

Model evaluation is used to measure how well the model can understand patterns in data, make accurate predictions, or perform specific tasks according to its purpose. Model evaluation is a critical step in model development and improvement, as it provides insight into potential improvements that can be made. Evaluation of the K-Means++ model using Sum of Squares (SSE), also known as Inertia, can provide an evaluation of how well the K-Means++ model separates data into different groups. SSE measures how far each data point is from a predetermined cluster center.

Table 4. Model Evaluation

Method	Value
Sum of Square (SSE)	0.6605

Based on Table 4, the SSE value of 0.6605 indicates that the data points tend to be spread wider than the cluster center, and the cluster may not be very compact. However, there are also scattered data points near the center of the cluster. This can be considered a fairly good indication of the cluster.

Conclusion

Based on the research that has been carried out, it can be concluded that the K-Means++ cluster analysis using the elbow method produces the optimal number of clusters k=3. Clustering based on health services showed that cluster 1 had 18 districts/cities, cluster 2 had 7 districts/cities, and cluster 3 had 2 districts/cities. Cluster three is a cluster dominated by developed districts/cities.

References

[1] S. Arifin, T. Lestarisa, R. A. A. H. S. P. S, A. Widiarti, D. Mutiasari, and H. Jelita, *Sistem Pelayanan Kesehatan Masyarakat*, 1st ed. Yogyakarta: CV Mine, 2022.

[2] Y. Pusparisa, "Bagaimana Kondisi Layanan Kesehatan di Indonesia?," *databoks*, 2020. https://databoks.katadata.co.id/datapublish/2020/04/03/bagaimana-kondisi-layanan-kesehatan-di-indonesia.

[3] A. Rahim, I. Pasari, and M. Sutanty, "Pengaruh Pendapatan Masyarakat Terhadap Permintaan Pelayanan Kesehatan di Kabupaten Sumbawa (Studi Pada RSUD Kabupaten Sumbawa)," *J. Ekon. Bisnis*, vol. 10, no. 3, pp. 304–312, 2022, doi: 10.58406/jeb.v10i3.1041.

[4] M. Misnaniarti *et al.*, "Ketersediaan Fasilitas dan Tenaga Kesehatan Dalam Mendukung Cakupan Semesta Jaminan Kesehatan Nasional," *J. Penelit. dan Pengemb. Pelayanan Kesehat.*, no. May 2019, pp. 6–16, 2018, doi: 10.22435/jpppk.v1i1.425.

[5] Yandrizal, D. Suryani, B. Anita, and H. Febriawati, "Analisis Ketersediaan Fasilitas Kesehatan dan Pemerataan Pelayanan Pada Pelaksanaan Jaminan Kesehatan Nasional di Kota Bengkulu, Kabupaten Seluma dan Kabupaten Saur," *J. Kebijak. Kesehat. Indones.*, vol. 03, no. 02, pp. 103–112, 2014.

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art11

[6] Z. A. Noor, T. D. Sekarningrum, and T. Sulistyaningsih, "Disparitas perkotaan-pedesaan: pemerataan dalam akses layanan kesehatan primer untuk lansia selama pandemi Covid-19," *JPPI (Jurnal Penelit. Pendidik. Indones.*, vol. 7, no. 4, p. 576, 2021, doi: 10.29210/020211249.

[7] U. F. Laili, C. Umatin, and M. U. Ridwanulloh, "Analisis Potensial Drop Out Mahasiswa Dengan K-Means++ Clustering Dalam Upaya Peningkatan Kualitas IAIN Kediri," *Paedagoria J. Kajian, Penelit. dan Pengemb. Kependidikan*, vol. 14, no. 2, pp. 145–153, 2023.

[8] A. R. Hayati, M. Hani'ah, and I. Kusumaning, "Comparison of Result Clustering Study Case Posyandu With The Scalable K Means ++ Clustering Method," *Conf. Senat. STT Adisutjipto Yogyakarta*, vol. 6, pp. 263–272, 2020, doi: 10.28989/senatik.v6i0.408.

[9] N. Nugroho and F. D. Adhinata, "Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa," *Teknika*, vol. 11, no. 3, pp. 170–179, 2022, doi: 10.34148/teknika.v11i3.502.

[10] C. M. Fikri, F. E. M. Agustin, and F. Mintarsih, "Pengelompokan Kualitas Kerja Pegawai Menggunakan Algoritma K-Means++ Dan Cop-Kmeans Untuk Merencanakan Program Pemeliharaan Kesehatan Pegawai Di Pt. Pln P2B Jb Depok," *Pseudocode*, vol. 4, no. 1, pp. 9–17, 2017, doi: 10.33369/pseudocode.4.1.9-17.

[11] R. Nainggolan, R. Perangin-Angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," *J. Phys. Conf. Ser.*, vol. 1361, no. 1, 2019, doi: 10.1088/1742-6596/1361/1/012015.