**Jurnal Statistika Dan Komputasi (STATKOM), Vol. 3 No. 2 (Dec 2024)**
**ISSN : 2963-038X (print), ISSN : 2963-0398**
**DOI :** https://doi.org/10.32665/statkom.v4i2.3592

# Clustering Study Of Hospitals In Bojonegoro Based On Health Workers With K-Means And K-Medoids Methods

Elsa Maulida Safitri[1]

[1]Statistics Study Program, Universitas Nahdlatul Ulama Sunan Giri
E-mail: elsafitri@unugiri.ac.id[1]

***Submitted*** 3 December 2024 ***Revised*** 25 December 2024 ***Accepted*** 30 December 2024

## *Abstract*

***Background:*** *Hospitals are institutions that provide inpatient care for the sick. In Bojonegoro, hospital services are considered adequate. However, a shortage of nurses often requires patients' families to assist with care.*
***Objective:*** *This research aims to compare clustering methods to find the best method that can be applied to cluster hospitals based on the type of health workers.*
***Methods:*** *This study uses two clustering methods, namely K-Means and K-Medoids Clustering, which are compared to determine the best method. The data source used is secondary data, which consists of the number of medical staff for each medical position, obtained from the Satu Data Bojonegoro website in 2020.*
***Results:*** *The K-means method proved to be the best for grouping healthcare workforce data. Its average within-cluster distance value is -6.763, the closest to zero. The K-means method resulted in 4 clusters. These include cluster_0 (3 hospitals), cluster_1 (1 hospital), cluster_2 (1 hospital), and cluster_3 (5 hospitals).*
***Conclusion:*** *The clustering results show that K-Means with 4 clusters is the best method, with Cluster_0 and Cluster_3 having below-average health workers, and Cluster_1 and Cluster_2 having above-average health workers, with Cluster_2 having the highest and Cluster_3 the lowest number of health workers in Bojonegoro.*

***Keywords :*** *Hospital, K-Means, K-Medoids.*

## Abstrak

**Latar Belakang**: Rumah sakit adalah institusi yang menyediakan perawatan rawat inap bagi orang sakit. Di Bojonegoro, pelayanan rumah sakit dinilai cukup baik. Namun, kekurangan tenaga perawat sering membuat keluarga pasien diminta membantu perawatan.
**Tujuan**: Penelitian ini bertujuan untuk membandingkan metode-metode clustering untuk menemukan metode terbaik yang dapat diterapkan untuk mengelompokkan rumah sakit berdasarkan jenis tenaga kesehatan.
**Metode:** Penelitian ini menggunakan dua metode clustering yaitu K-Means dan K-Medoids Clustering yang dibandingkan untuk mendapatkan metode terbaik. Sumber data yang digunakan adalah data sekunder yang merupakan banyaknya tenaga kerja medis untuk setiap posisi tenaga medis yang diambil dari website Satu Data Bojonegoro pada tahun 2020.
**Hasil:** Metode K-means terbukti menjadi metode terbaik untuk mengelompokkan data jumlah tenaga kesehatan. Nilai average within cluster distance dari metode ini adalah -6,763, yang paling mendekati nol. Metode K-means menghasilkan 4 cluster. Cluster tersebut meliputi cluster_0 (3 rumah sakit), cluster_1 (1 rumah sakit), cluster_2 (1 rumah sakit), dan cluster_3 (5 rumah sakit).
**Kesimpulan:** Hasil pengelompokan menunjukkan bahwa metode K-Means dengan 4 cluster adalah yang terbaik, di mana Cluster_0 dan Cluster_3 memiliki jumlah tenaga kesehatan di bawah rata-rata, sedangkan Cluster_1 dan Cluster_2 memiliki jumlah tenaga kesehatan di atas rata-rata, dengan Cluster_2 memiliki jumlah tertinggi dan Cluster_3 jumlah terendah di Bojonegoro.

**Kata kunci:** Rumah Sakit, K-Means, K-Medoids.

## INTRODUCTION

A hospital is a health agency and health center that accommodates and treats sick people and provides inpatient care (Hutagalung, 2022). The allocation of government funds can support the need for quality health services, so the hospital needs to maintain employee performance so that its goals are achieved in providing quality health services to the community, where the performance of health workers dramatically affects the quality of hospital services (Rachdiansyah & Tesmanto, 2021).

According to data collected from the number of health workers in Bojonegoro Regency, hospital patient care is quite good, but there is one obstacle. The shortage of nurses causes health workers to sometimes ask the patient's family for help (Khayudin et al., 2022). A more significant number of health workers can reduce the number of disease cases in the next period (Nurdiansyah & Sulistiawan, 2023).

Considering these issues, each hospital requires adequate medical personnel. For this problem, hospitals in Bojonegoro should be grouped according to the type of health workers. The goal is to improve the efficiency of each hospital's health personnel placement policy by providing knowledge to the relevant agencies. The purpose of the study is to classify hospitals based on the type of health workers in Bojonegoro. In this case, the appropriate statistical method is clustering, discussed in the data mining study.

Pattern recognition is decreasing as it has become part of data mining. This method combines four fields: statistics, visualization, databases, and machine learning. Machine learning is a field of artificial intelligence that is related to developing programming techniques based on prior learning and sometimes related to statistics and optimization (Mujiasih, 2011).

Data mining for weather prediction has been widely used. The selection of data mining techniques using association rules with the Apriori algorithm has shown better results regarding correctness, computational process, and termination (Nandagopal et al., 2010). Data mining is an analytical process for discovering knowledge in large and complex data sets at the intersection of statistics and computer science. Furthermore, data mining combines statistics, computer science, artificial intelligence, and machine learning (Buaton et al., 2020). Data mining can be divided into several groups: prediction methods, cluster analysis, association analysis, and anomaly detection (Sibuea & Safta, 2017).

Clustering impacts data collection through recording, correspondence, or analyzing and organizing groups of objects with similar characteristics. A cluster is a collection of records that have identical characteristics to each other and are different from records in other clusters. Clustering can be used to group all data into relatively small groups with similarities; in this case, the similarity between groups will be the largest, while the differences between groups will be the most minor (Zulfa & Hadiana, 2021). Data clustering is done by equalizing the characteristics of each data set so that the label corresponding to the group formed can be determined. This results in the clustering method, also known as segmentation (Zulfa & Hadiana, 2021). The following will present an example of a 3-group clustering image in Figure 1.

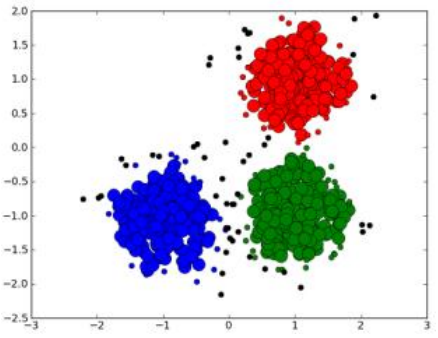

Figure 1. Clustering of the Three Groups

Source: Journal of RapidMiner Studio 8.2 Utilization for Clustering Accessories Sales Data Using K-Means Algorithm (Mardalius, 2018).

The K-Means algorithm partitions the database into k groups and is the most popular classifier method. It focuses on forming cluster partitions using the centroid or mean of a set of data. The algorithm starts with the formation of an initial cluster partition and iteratively refines it until there is no significant change.

The Elbow method is used to determine the optimal number of clusters by looking at the percentage change in the Sum of Squared Errors (SSE) at a certain point (Fitriyah et al., 2023). SSE is calculated for each cluster value, and a decrease in SSE indicates an increase in cluster homogeneity. The steps of the Elbow method in determining the value of k in the K-Means algorithm are as follows. First, giving initial initialization of k value. Second, increasing the value of k gradually. Third, calculating the Sum of the Squared Error (SSE) of each k value. Fourth, observing the SSE results, looking for the point where the decrease in SSE becomes more gentle. Fifth, determining the value of k at which the SSE graph forms an elbow.

K-Medoids, also known as the Partitioning Around Medoids (PAM) algorithm, is an algorithm that uses medoids as the main representation of each cluster (Kamila et al., 2019). The K-Medoids method distinguishes itself from other methods by using medoids as cluster centers that represent other members, in contrast to K-Means, which uses the average value as the cluster center. The partitional clustering technique in K-Medoids involves minimizing the distance between a point and the medoid chosen as the cluster center.

K-Medoids cluster datasets use the classic partitioning technique in clustering, where data is grouped into k groups, also known as a priori. K-Medoids' advantage over other methods is their efficiency in handling noise and outliers in the data. It is due to dissimilarity, calculated as the depreciation of the distance between the object and the medoid rather than the sum of the squares of the Euclidean distances. This method can also be known as the average dissimilarity over all objects in the cluster, with the medoid as the object that best represents the cluster (Anggreini & Tresnawati, 2020). The steps of the K-Medoids method can be explained as follows. First, giving initialize k cluster centers.Second, allocating each object to the nearest cluster using the Euclidean distance formula:

$$d(x,y) = \|x - y\|$$

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \;\; ; \;\; i = 1,2,3,\dots,n$$

Third, randomly selecting objects from each cluster and assign them as new medoid candidates. Fourth, calculating the distance from all objects in the cluster to the new medoid candidate. Fifth, calculating the total deviation (S) by subtracting the previous total distance from the new total distance and, if necessary, replacing objects in the cluster with the new medoid. Sixth, repeating steps 3 to 5 until no significant changes occur in the medoids and until stable members are obtained for each cluster.

The most common data mining method is clustering, where data points are grouped into two or more based on an ideal centroid point. Each data point belongs to a group with similar characteristics to each other based on the available information compared to a different group (Herlinda et al., 2021). Using the RapidMiner application becomes easier in studies that use clustering methods. In addition to the K-Means method, another clustering method used is K-Medoids. This method was chosen because it can overcome the shortcomings of clustering methods, especially K-Means, including handling outlier and noise data (Sholikhah, 2022). K-Medoids have been applied in various fields, such as clustering immunization data in Indonesia (Sundari et al., 2019), clustering historical sites in Malang City based on the web, and clustering loading and unloading transaction data in Riau Province in the field of

transportation (Kamila et al., 2019).

This study used descriptive statistics tables to perform statistical descriptions. The novelty is that the health worker data from each hospital in Bojonegoro is divided into several types of jobs. These hospitals are grouped based on these types of health workers by comparing two clustering methods, namely the K-means and K-Medoids methods.

This research aims to obtain descriptive statistics of hospital health workers in Bojonegoro Regency in 2020 and compare the K-means and K-Medoids methods in clustering hospitals by type of health worker. The analysis results will determine the best clustering method for optimal grouping. Thus, this research is expected to provide valuable recommendations for managing health workers in hospitals in the area.

The benefits obtained from this research are that the Bojonegoro Regency Health Office can obtain consideration and knowledge in taking the next step and in making policies related to the procurement and distribution of health workers in Bojonegoro Regency. The need for and placement of health workers in each hospital is expected to be by the needs and be efficient. It is done so that disease management can be more prepared and alert and that hospitals can anticipate unpredictable labor needs. Therefore, a research called "Clustering Study Of Hospitals In Bojonegoro Based On Health Workers With K-Means And K-Medoids Methods" should be proposed.

## METHOD

### Research Design

This study employs a quantitative research design with an observational approach, utilizing the K-Means and K-Medoids methods to cluster hospitals in Bojonegoro based on the distribution of health workers. K-Means partitions the data into k clusters by minimizing the squared distances between data points and cluster centroids, while K-Medoids, being more robust to outliers, uses medoids as cluster centers.

### Population and Sample

The population in this study consists of all hospitals in Bojonegoro Regency along with the health workers employed in those hospitals. The sample observed includes hospitals that provide complete and accessible data on the distribution and characteristics of their health workers. This sample is considered representative of the population as it covers hospitals of various types (e.g., public, private, general, and specialty hospitals) within the region.

### Sampling Techniques

The study uses a purposive sampling technique, where samples are selected based on specific criteria, such as the availability, completeness, and reliability of data related to health workers. This technique ensures that the sample accurately represents the population and provides meaningful insights for the clustering analysis.

### Research Subjects

The data source used in this research is secondary data in the form of health worker data from the Satu Data Bojonegoro website. Observation data, which included ten hospital units in Bojonegoro Regency, was recorded in the database in 2020. It are RSUD Sumberrejo, RSUD Sosrodoro Djatikoesoemo, RSUD Padangan, RSIM Sumberrejo, RSIM Kalitidu, RSIA Fatma, RS Muna Anggita, RS Ibnu Sina, RS Bhayangkara Wahyu Tetuko, and RS 'Aysiyah. The research variables used in this study will be given their definitions and presented in Table 1 as follows.

Tabel 1. Definition of Research Variables

| Number | Variable | Measurement Scale | Description |
|---|---|---|---|
| 1 | Nurse | Ratio | Number of health workers for nurses |
| 2 | Support | Ratio | Number of health workers for support |
| 3 | Specialist Doctor | Ratio | Number of health workers for specialist doctors |
| 4 | Midwife | Ratio | Number of health workers for midwives |
| 5 | Engineering | Ratio | Number of health workers for engineering |
| 6 | General Physician | Ratio | Number of health workers for general practitioners |
| 7 | Pharmaceutical Technical Personnel | Ratio | Number of health workers for technical personnel |
| 8 | Pharmacist | Ratio | Number of health workers for pharmacists |
| 9 | Nutrition | Ratio | Number of health workers for nutrition |
| 10 | Physical Therapist | Ratio | Number of health workers for physical therapy |
| 11 | Environmental Health | Ratio | Number of health workers for environmental health |
| 12 | Dentist | Ratio | Number of health workers for dentists |
| 13 | Specialist Dentist | Ratio | Number of health workers for specialist dentists |
| 14 | Public Health | Ratio | Number of health workers for public health |

The data structure in the method used in the dataset follows the data structure in Table 2 below.

Table 2. Data Structure for Research

| Unit | Nurse | Support | ... | Public Health |
|---|---|---|---|---|
| RSUD Sumberrejo | $X_{1,1}$ | $X_{1,2}$ | ... | $X_{1,14}$ |
| RSUD Sosrodoro Djatikoesoemo | $X_{2,1}$ | $X_{2,2}$ | ... | $X_{2,14}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| RS 'Aysiyah | $X_{10,1}$ | $X_{10,2}$ | ... | $X_{10,14}$ |



## Data Analysis Techniques

The next stage includes evaluating the clustering method with the following steps (Sholikhah, 2022).

1. Input of observation data

At this stage, hospital data is entered into the RapidMiner software. The dataset input is in the form of a file with the .xls extension containing health worker data.

2. Retrieve observation data

After the data has been inputted with the appropriate data type, it can be entered into the analysis process by dragging the dataset into the designed analysis process. The following is a display of the dataset dragged from the file with the .xls extension, as seen in Figure 2 for the display of retrieved data used in the clustering method.

3. Method evaluation

3.1. The K-Means method is evaluated using K-means clustering tools for performance testing. Performance testing can be done using cluster density performance tools. There are 4 clusters to see the cluster results of the data used in this K-means clustering algorithm. The following is Figure 2, which displays the data processing process using the K-Means algorithm.

3.2. The K-Medoids method is evaluated using the K-Medoids clustering tools operator for performance testing, using cluster density performance tools to find the cluster results from the data. The cluster parameter used in this K-Medoids algorithm is 2. Figure 2 shows the data clustering process with the K-Medoids algorithm.

4. Performance Vector

The performance vector used for this research is cluster density performance,

which produces a good average value within clusters (average within cluster distance value) for the K-Means and K-Medoids methods. The following is a view of Figure 2 regarding the design of using the performance vector in the clustering method comparison process.

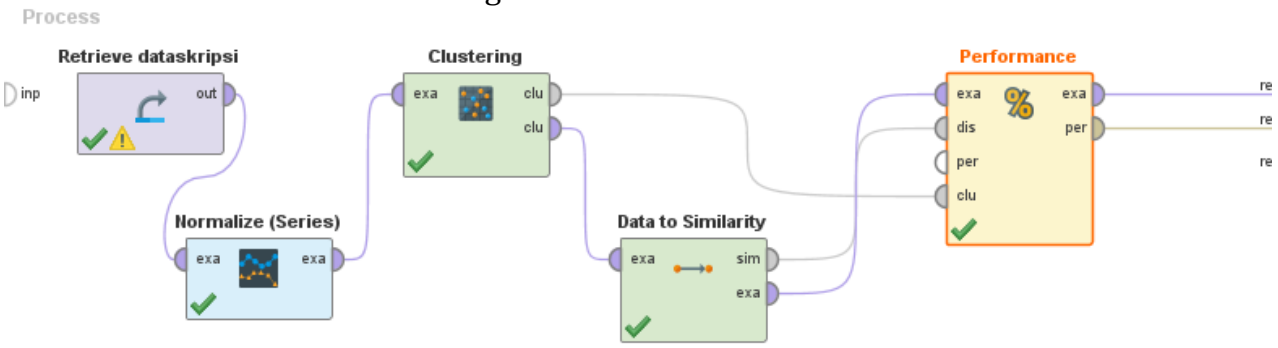


Figure 2. Display Design Process

The clustering method comparison process compares the performance vector value with the best cluster k based on the centroid-based average within the cluster value. After comparing clustering methods, the best clustering model for clustering hospitals in Bojonegoro Regency will be determined. Hospital clustering based on the type of health worker is done by applying the best clustering method.

## RESULTS AND DISCUSSION

### Descriptive Statistics

The observation data shows that the number of samples is ten hospitals with 15 variables representing various types of health workers in each hospital. The following is a summary of the descriptive statistics table for the observed research variables in Table 3.

Table 3. Summary of Descriptive Statistics

| Variable | *n* | Mean | Standart Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Nurse | 10 | 82 | 68.321 | 31 | 247 |
| Support | 10 | 62 | 48.287 | 0 | 174 |
| Specialist Doctors | 10 | 17 | 12.665 | 6 | 42 |
| Midwife | 10 | 22 | 13.47 | 5 | 46 |
| Engineering | 10 | 5 | 6.999 | 0 | 23 |
| General Physician | 10 | 9 | 4.572 | 6 | 20 |
| Pharmaceutical Technical Personnel | 10 | 4 | 4.858 | 0 | 15 |
| Pharmacist | 10 | 3 | 3.206 | 0 | 10 |
| Nutrition | 10 | 2 | 2.716 | 0 | 9 |
| Physical Therapist | 10 | 1 | 2.014 | 0 | 6 |
| Environmental Health | 10 | 0 | 1.595 | 0 | 5 |
| Dentist | 10 | 1 | 0.675 | 0 | 2 |
| Specialist Dentist | 10 | 0 | 0.632 | 0 | 2 |
| Public Health | 10 | 1 | 1.932 | 0 | 6 |

In 2020, the number of nurse in Bojonegoro Regency hospitals had an average of 82 with a standard deviation of 68.321 , with the lowest number of workers being 31 and the highest being 247. Supporting variables have an average of 62 with a standard deviation of 48.287, a minimum number of 0 and a maximum of 174. The average number of midwives is 22, with a standard deviation of 13.47, a minimum of 5, and a maximum of 46. Specialist doctors have an average of 17 with a standard deviation of 12.665, a minimum of 6, and a maximum of 42. The average number of general practitioners is 9, with a standard deviation of 4.572, a

minimum of 0, and a maximum of 20. The average number of technical personnel is 5, with a standard deviation of 6.999, a minimum of 0, and a maximum of 23. Pharmaceutical technical personnel have an average of 4 with a standard deviation of 4.858, a minimum of 0, and a maximum of 10.

The average number of pharmacists is 3, with a standard deviation of 3.206, a minimum of 0, and a maximum of 10. The number of nutrition workers averages 2 with a standard deviation of 2.716, a minimum of 0, and a maximum of 6. Dentists have an average of 1 with a standard deviation of 0.67, a minimum of 0, and a maximum of 2. Environmental health workers have an average of 0 with a standard deviation of 1.595, a minimum of 0, and a maximum of 5. Physical therapists have an average of 1 with a standard deviation of 2.04, a minimum of 0, and a maximum of 6. Dental specialists have an average of 0 with a standard deviation of 0.632, a minimum of 0, and a maximum of 2. Public health workers have an average of 1 with a standard deviation of 1.932, a minimum of 0, and a maximum of 6.

## Hospital Grouping in Bojonegoro Regency

This research has carried out the performance evaluation of the K-means vector. Table 4 presents in detail the testing of each cluster with the value of k in the K-means method.

Table 4. Investigation Study of K-Means Method

| Grouping | Average Within Cluster Distance | Difference |
|---|---|---|
| K = 2 | -18.256 | - |
| K = 3 | -16.696 | +1.56 |
| K = 4 | **-6.763** | +9.933 |
| K = 5 | -4.213 | +2.55 |
| K = 6 | -3.003 | +1.21 |
| K = 7 | -1.495 | +1.508 |
| K = 8 | -0.711 | +0.784 |
| K = 9 | -0.311 | +0.4 |
| K = 10 | 0.00 | +0.311 |

From Table 4, it is known that the average value of within-cluster distance on the performance vector of each group with the value of k = 2 is -18.256, the value of k = 3 is -16.696, the value of k = 4 is -11.173 until the lowest value at k = 10 is 0. In conclusion, the best k selection in the K-means method is 4 (first slope position) based on Fitriyah et al. (2023). The following is presented in Table 5, which shows the results of grouping health workers in Bojonegoro hospitals with the K-Means method.

Table 5. Hospital Grouping Results with K-Means Method

| K-Means | Unit | X1 | X2 | X3 | X4 |
|---|---|---|---|---|---|
| cluster_0 | RS Sumberrejo, RS Padangan, dan RSIA Fatma | 67.00 | 61.67 | 34.67 | 11.67 |
| cluster_1 | RS 'Aisyiyah | 77.00 | 152.00 | 13.00 | 39.00 |
| cluster_2 | RSUD Sosrodoro Djatikoesoemo | 174.00 | 247.00 | 42.00 | 42.00 |
| cluster_3 | RSIM Sumberrejo, RSIM Kalitidu, RS Muna Anggita, RS Ibnu Sina, dan RS Bhayangkara Wahyu Tetuko | 34.60 | 47.20 | 14.00 | 12.40 |

Continued Table 5. Hospital Grouping Results with K-Means Method

| K-Means | Unit | X5 | X6 | X7 | X8 | X9 |
|---|---|---|---|---|---|---|
| cluster_0 | RS Sumberrejo, RSPadangan, dan RSIAFatma | 7.00 | 4.67 | 4.33 | 3.00 | 2.67 |
| cluster_1 | RS 'Aisyiyah | 20.00 | 2.00 | 9.00 | 8.00 | 1.00 |
| cluster_2 | RSUD Sosrodoro Djatikoesoemo | 16.00 | 23.00 | 15.00 | 10.00 | 9.00 |
| cluster_3 | RSIM Sumberrejo, RSIM Kalitidu, RS Muna Anggita, RS Ibnu Sina, dan RS Bhayangkara Wahyu Tetuko | 8.00 | 4.00 | 1.80 | 1.60 | 1.20 |

Continued Table 5. Hospital Grouping Results with K-Means Method

| K-Means | Unit | X10 | X11 | X12 | X13 | X14 |
|---|---|---|---|---|---|---|
| cluster_0 | RS Sumberrejo, RS Padangan, dan RSIA Fatma | 0.67 | 1.00 | 1.67 | 0.00 | 0.33 |
| cluster_1 | RS 'Aisyiyah | 2.00 | 1.00 | 2.00 | 0.00 | 6.00 |
| cluster_2 | RSUD SosrodoroDjatikoesoemo | 2.00 | 5.00 | 6.00 | 2.00 | 1.00 |
| cluster_3 | RSIM Sumberrejo,RSIM Kalitidu, RS Muna Anggita, RS Ibnu Sina, dan RS Bhayangkara Wahyu Tetuko | 1.40 | 0.00 | 0.40 | 0.00 | 0.80 |

After conducting research using the first method, the research continued with the second method, namely K-Medoids. Table 6 shows each cluster's detailed testing with the k value in the K-Medoids method.

Table 6. Investigation Study of K-Medoids Method

| Grouping | Average Within Cluster Distance | Difference |
|---|---|---|
| K = 2 | -18.258 | - |
| K = 3 | -16.696 | +1.562 |
| K = 4 | -12.073 | +4.623 |
| K = 5 | **-7.377** | +4.696 |
| K = 6 | -4.272 | +3.105 |
| K = 7 | -2.475 | +1.797 |
| K = 8 | -1.058 | +1.417 |
| K = 9 | -0.311 | +0.747 |
| K = 10 | 0.00 | +0.311 |

From Table 6, it is known that the average value of within-cluster distance on the performance vector of each group with the value of k = 2 is -18.256, the value of k = 3 is -16.696, the value of k = 4 is -12.073 until the lowest value at k = 10 is 0. In conclusion, the best k selection in the K-means method is 5 (first slope position) based on Fitriyah et al. (2023). The following is presented in Table 7, which shows the results of grouping health workers in Bojonegoro hospitals with the K-Medoids method.

Table 7. Hospital Grouping Results with K-Medoids Method

| K- Medoids | Unit | X1 | X2 | X3 | X4 |
|---|---|---|---|---|---|
| cluster_0 | RSUD Sosrodoro Djatikoesoemo | 174.00 | 247.00 | 42.00 | 42.00 |
| cluster_1 | RS Padangan | 96.00 | 94.00 | 28.00 | 15.00 |
| cluster_2 | RS Sumberrejo, RSIMSumberrejo, RSIMKalitidu, RSIA Fatma,RS Muna Anggita, dan RS Ibnu Sina | 33.67 | 46.33 | 22.50 | 10.33 |
| cluster_3 | RS Bhayangkara Wahyu Tetuko | 76.00 | 49.00 | 11.00 | 20.00 |
| cluster_4 | RS 'Aisyiyah | 77.00 | 152.00 | 13.00 | 39.00 |

Continued Table 7. Hospital Grouping Results with K-Means Method

| K-Medoids | Unit | X5 | X6 | X7 | X8 | X9 |
|---|---|---|---|---|---|---|
| cluster_0 | RSUD Sosrodoro Djatikoesoemo | 16.00 | 23.00 | 15.00 | 10.00 | 9.00 |
| cluster_1 | RS Padangan | 7.00 | 9.00 | 8.00 | 5.00 | 5.00 |
| cluster_2 | RS Sumberrejo, RSIM Sumberrejo, RSIM Kalitidu, RSIA Fatma, RS Muna Anggita, dan RS Ibnu Sina | 7.50 | 4.17 | 2.33 | 1.67 | 1.00 |
| cluster_3 | RS Bhayangkara Wahyu Tetuko | 9.00 | 0.00 | 0.00 | 2.00 | 3.00 |
| cluster_4 | RS 'Aisyiyah | 20.00 | 2.00 | 9.00 | 8.00 | 1.00 |

Continued Table 7. Hospital Grouping Results with K-Means Method

| K-Medoids | Unit | X10 | X11 | X12 | X13 | X14 |
|---|---|---|---|---|---|---|
| cluster_0 | RSUD Sosrodoro Djatikoesoemo | 2.00 | 5.00 | 6.00 | 2.00 | 1.00 |
| cluster_1 | RS Padangan | 1.00 | 2.00 | 4.00 | 0.00 | 0.00 |
| cluster_2 | RS Sumberrejo, RSIMSumberrejo, RSIMKalitidu, RSIA Fatma,RS Muna Anggita, dan RS Ibnu Sina | 1.00 | 0.17 | 0.33 | 0.00 | 0.33 |
| cluster_3 | RS Bhayangkara Wahyu Tetuko | 2.00 | 0.00 | 1.00 | 0.00 | 3.00 |
| cluster_4 | RS 'Aisyiyah | 2.00 | 1.00 | 2.00 | 0.00 | 6.00 |

## Comparison Process of Clustering Methods

In the comparison process, we will compare the performance vector's value with each clustering method's best results. Based on the clustering analysis test results of the centroid-based K-means and K-Medoids methods, the comparison results of the two clustering methods are described in Figure 3 as follows.

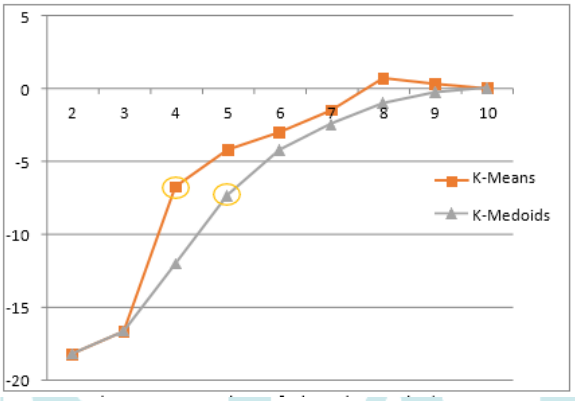


Figure 3. Comparison of Clustering Methods

Based on Figure 3 above, the best clustering method is selected based on the Average Within Cluster Distance value close to zero, which is at k = 4 with the K-means method. Further explanation can be found in Table 4.6 regarding comparing clustering methods.

Table 8. Comparison of Clustering Methods

| | Average Within Cluster Distance | |
|---|---|---|
| Grouping (k) | K-Means | K-Medoids |
| 2 | -18.256 | -18.258 |
| 3 | -16.696 | -16.696 |
| 4 | **-6.763** | -12.073 |
| 5 | -4.213 | **-7.377** |
| 6 | -3.003 | -4.272 |
| 7 | -1.495 | -2.475 |
| 8 | -0.711 | -1.058 |
| 9 | -0.311 | -0.311 |
| 10 | 0.00 | 0.00 |

Table 8 shows the Average Within Cluster Distance value from the 2nd to the 10th cluster with the best K-means method. The first stable value selected from K-means is at cluster k = 4 with a value of - 6.763, while from K-Medoids, the first stable value is at cluster k = 5 with a value of -7.377. The best method is the clustering method with the Average Within Cluster Distance value close to zero, namely the K-Means Method at k of 4.

## CONCLUSION

### Conclusion

Based on descriptive statistical analysis, the lowest number of health workers in Bojonegoro Regency is 52 at Muna Anggita Bojonegoro Hospital, while the highest number is 594 at Sosrodoro Djatikoesoemo Bojonegoro Hospital, with a mean value of 216 and a standard deviation of 154.542. A comparison of the K-means method shows that the first stable point in the average within distance value is seen at k = 4 with a value of -6.763, which then stabilizes at k = 5 with a value of -5.341.

The clustering results indicate that the best method used is K-Means with 4 clusters. First, Cluster_0 and Cluster_3 represent groups of hospitals with a below-average number of health workers in Bojonegoro, where Cluster_3 has the lowest number of health workers, followed by Cluster_0, which ranks second-lowest. Second, Cluster_1 and Cluster_2 represent groups of hospitals with an above-average number of health workers in Bojonegoro, where Cluster_2 has the highest number of health workers, followed by Cluster_1, which ranks second-highest.

### Recommendations

Based on the results of this study, it is recommended that hospital managers in Bojonegoro Regency consider using the K-means method in analyzing and clustering health worker data for better planning. Hospitals with much fewer health workers need special attention to ensure optimal service availability. In addition, it is recommended that further research be

conducted by including other relevant variables to obtain a more comprehensive picture of the distribution and needs of health workers in the region.

## BIBLIOGRAPHY


Anggreini, N. L., & Tresnawati, S. (2020). Komparasi Algoritma K-Means Dan K-Medoids Untuk Menangani Strategi Promosi Di Politeknik TEDC Bandung. Jurnal TEDC, 14(2), 120–127.

Buaton, R., Zarlis, M., & Mawengkang, H. (2020). Model Optimasi Prediksi dengan Model Association Rule Best Time Series (ARBT) Pada Data Mining Time Series. ... Teknologi Komputer & ..., 715–720. https://prosiding.seminar-id.com/index.php/sainteks/article/view/538

Fitriyah, H., Safitri, E. M., Muna, N., Khasanah, M., Aprilia, D. A., & Nurdiansyah, D. (2023). IMPLEMENTASI ALGORITMA CLUSTERING DENGAN MODIFIKASI METODE ELBOW UNTUK MENDUKUNG STRATEGI PEMERATAAN BANTUAN SOSIAL DI KABUPATEN BOJONEGORO. Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika, 4(3), 1598–1607. https://doi.org/https://doi.org/10.46306/lb.v4i3.453

Herlinda, V., Darwis, D., & Dartono, D. (2021). ANALISIS CLUSTERING UNTUK RECREDESIALING FASILITAS KESEHATAN MENGGUNAKAN METODE FUZZY C-MEANS. Jurnal Teknologi Dan Sistem Informasi, 2(2), 94–99. https://doi.org/https://doi.org/10.33365/jtsi.v2i2.890

Hutagalung, L. E. (2022). Analisa Manajemen Risiko Sistem Informasi Manajemen Rumah Sakit (Simrs) Pada Rumah Sakit Xyz Menggunakan Iso 31000. Jurnal Telka, 12(01), 23–33. https://doi.org/10.36342/teika.v12i01.2820

Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi, 5(1), 119–125.

Khayudin, B. A., Nurfain, & Retno Kusuma Hati, D. (2022). PENGALAMAN PERAWAT DALAM MERAWAT PASIEN TOTAL CARE DI RUANG ICU RSUD DR. R. SOSODORO DJATIKOESOEMO BOJONEGORO. Jurnal Ilmu Kesehatan MAKIA, 12(2), 111–118. https://doi.org/10.37413/.v12i2.235

Mardalius. (2018). Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. IV(2), 401–411.

Mujiasih, S. (2011). Pemanfatan Data Mining Untuk Prakiraan Cuaca. Jurnal Meteorologi Dan Geofisika, 12(2), 189–195. https://doi.org/10.31172/jmg.v12i2.100

Nandagopal, S., Karthik, S., & Arunachalam, V. P. (2010). Mining of meteorological data using Modified Apriori algorithm. European Journal of Scientific Research, 47(2), 295–308.

Nurdiansyah, D., & Sulistiawan, A. (2023). PEMODELAN JUMLAH KASUS DEMAM BERDARAH DENGUE DENGAN MENGGUNAKAN MODEL AUTOREGRESSIVE DISTRIBUTED LAG. Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika, 4(3), 1965–1977. https://doi.org/10.46306/lb.v4i3.526

Rachdiansyah, I., & Tesmanto, J. (2021). Pengaruh Audit Manajemen Sumber

Daya Manusia terhadap Kinerja Karyawan di Rumah Sakit Umum Daerah Kota Bekasi. VISA: Journal of Vision and Ideas, 1(1), 1–13. https://doi.org/10.47467/visa.v1i1.756

Sholikhah, N. A. (2022). Studi Perbandingan Clustering Kecamatan di Kabupaten Bojonegoro Berdasarkan Keaktifan Penduduk Dalam Kepemilikan Dokumen Kependudukan. Jurnal Statistika Dan Komputasi, 1(1), 42–53. https://doi.org/10.32665/statkom.v1i1.443

Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring. JURTEKSI, 4(1), 85–92. https://doi.org/10.33330/jurteksi.v4i1.28

Sundari, S., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, J., & Wanto, A. (2019). Analisis K-Medoids Clustering Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia. Prosiding Seminar Nasional Riset Information Science (SENARIS), 687–696. https://doi.org/10.30645/senaris.v1i0.75

Zulfa, N. S. L., & Hadiana, A. (2021). KAJIAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS DAN K-MEDOIDS DALAM STRATEGI PROMOSI (Studi Kasus: Universitas Islam Al-Ihya Kuningan) Neng. Jurnal Fakultas Teknik, 2(2), 57–62. https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part