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.v4i2.3532

STATKOM
JURNAL STATISTIKA DAN KOMPUTASI

Comparison of K-Means and Fuzzy C-Means for Optimizing Tuberculosis Management and Healthcare Service Allocation in Bojonegoro

Riko Al Muiz[1]

[1]Customer Service and Funds, PD BPR Bank Daerah Bojonegoro
E-mail: ikomulyo8@gmail.com[1]

Submitted 3 December 2024 ***Revised*** 16 December 2024 ***Accepted*** 20 December 2024

Abstract

Background: *According to the 2022 publication by BPS (Statistics Bureau) of Bojonegoro Regency, there were 1,765 tuberculosis cases spread across all districts in Bojonegoro. This number is disproportionate to the availability of healthcare workers, which totaled only 1,261, comprising medical personnel, nurses, midwives, and pharmacists.*
Objective: *This study aims to cluster districts in Bojonegoro Regency based on tuberculosis cases and healthcare workforce data by comparing the K-Means and Fuzzy C-Means methods. The objective is to identify which districts require more attention and which are already in better condition.*
Methods: *The best clustering method was determined using the Sum of Squared Error (SSE) criterion. The data used in this study was sourced from the Statistics Bureau, containing information on tuberculosis cases and the number of healthcare workers in each district..*
Results: *The result shows that K-Means achieved a lower SSE (4704.031) compared to Fuzzy C-Means (4854.247), which divided the district into 4 clusters: low, medium, and high. By categorizing the districts into these clusters, the Bojonegoro government is expected to better target its interventions and resources. Moreover, the government can evaluate districts with high tuberculosis cases to implement specific strategies.*
Conclusion: *This study concludes that K-Means with 4 clusters is the most effective method for this type of clustering.*

Keywords : *Clustering, K-Means, Fuzzy C-Means.*

Abstrak

Latar Belakang: Menurut data publikasi BPS Kabupaten Bojonegoro tahun 2022, terdapat 1.765 penderita tuberkulosis yang tersebar di seluruh kecamatan di Kabupaten Bojonegoro. Jumlah penderita ini tidak sebanding dengan ketersediaan tenaga kesehatan, yang hanya berjumlah 1.261 orang. Tenaga kesehatan tersebut terdiri dari tenaga medis, perawat, bidan, dan apoteker.
Tujuan: . Penelitian ini bertujuan untuk mengelompokkan kecamatan di Kabupaten Bojonegoro berdasarkan data kasus tuberkulosis dan jumlah tenaga kesehatan dengan membandingkan metode K-Means dan Fuzzy C-Means. Tujuannya adalah untuk mengidentifikasi kecamatan yang membutuhkan perhatian lebih serta kecamatan yang sudah memiliki kondisi yang memadai.
Metode: Pemilihan metode terbaik dilakukan berdasarkan nilai Sum of Squared Error (SSE). Sumber data yang digunakan berasal dari Badan Pusat Statistik, yang mencakup informasi kasus tuberkulosis dan jumlah tenaga kesehatan di setiap kecamatan..
Hasil: Hasilnya menunjukkan bahwa K-Means mencapai SSE yang lebih rendah (4704.031) dibandingkan dengan Fuzzy C-Means (4854.247), yang membagi kabupaten ke dalam 4 klaster yaitu rendah, sedang, dan tinggi. Pembagian data ke dalam kluster ini diharapkan dapat membantu pemerintah Kabupaten Bojonegoro untuk melakukan pengelompokan wilayah dengan lebih efektif. Selain itu, pemerintah juga dapat mengevaluasi wilayah yang memiliki jumlah kasus tinggi untuk dilakukan intervensi lebih lanjut.
Kesimpulan: Penelitian ini menyimpulkan bahwa metode K-Means dengan 4 kluster adalah metode yang paling efektif untuk pengelompokan data tersebut.

Kata kunci: Pengelompokkan, K-Means, Fuzzy C-Means.

INTRODUCTION

Chronic diseases are long-term conditions or illnesses, such as HIV/AIDS, dengue fever (DF), sexually transmitted infections (STIs), diarrhea, tuberculosis (TB), and malaria. Accelerating the recovery process for chronic disease patients requires effective collaboration between patients and healthcare providers. These providers include medical personnel, clinical psychologists, nurses, midwives, and pharmacists (Jaya & Oktafianto, 2022).

According to Mar'iyah and Zulkarnain (2021), tuberculosis is classified as a disease caused by Mycobacterium tuberculosis, a robust gram-positive bacterium that primarily attacks organs through the lungs. Tuberculosis often occurs in individuals with positive AFB (Acid-Fast Bacilli) and is transmitted via tiny droplets expelled when a patient coughs or sneezes. The pathophysiology of tuberculosis begins when the microbes enter the alveoli and provoke a response from the immune system. Tuberculosis transmission can be influenced by various factors, including age, lifestyle, smoking habits, irregular work patterns, and environmental conditions. Patients with tuberculosis generally exhibit symptoms such as a persistent cough lasting more than 14 days, shortness of breath, fatigue, loss of appetite, blood-streaked phlegm, fever, and weight loss (Sejati & Sofiana, 2015).

Tuberculosis cases in Indonesia have shown a significant increase over the past three years. The government needs to pay greater attention and implement effective preventive measures to reduce the number of cases. In 2015, 330,910 tuberculosis cases were reported, increasing to 360,565 cases in 2016 and 425,089 cases in 2017. According to a review of tuberculosis prevalence in 2013-2014, the prevalence of bacteriologically confirmed tuberculosis in Indonesia was 759 per 100,000 people aged 15 years and older, while the prevalence of AFB-positive TB was 257 per 100,000 people aged 15 years and older (Mathofani & Febriyanti, 2020).

According to the 2022 publication by BPS Bojonegoro Regency, there are 1,261 healthcare workers distributed across all districts in the region. Meanwhile, the number of residents infected with chronic diseases, including tuberculosis, remains high in almost every district. This infectious disease not only disrupts daily activities but also has the potential to increase mortality rates in Bojonegoro Regency. To address chronic disease cases, greater attention from both the government and the community is needed to achieve better public health outcomes. One crucial step is clustering districts based on tuberculosis cases and the number of healthcare workers. This clustering effort can support initiatives to reduce tuberculosis cases and improve the overall well-being of Bojonegoro's residents. Clustering is a research method used to group information into clusters so that data within the same cluster shares significant similarities (Hidayat et al., 2017). Clustering is also referred to as a part of data mining, aimed at enriching information and generating new data (R. D. Ramadhani, 2014). This technique helps minimize variations between data points and ensures that the objectives of the clustering system are met efficiently (Anggraini & Muharom, 2017).

Essentially, clustering is divided into two strategies: hierarchical and non-hierarchical methods. The hierarchical algorithm identifies patterns sequentially, where the groups have already been defined or chosen beforehand (L. Ramadhani et al., 2018). Data mining refers to the process of collecting or investigating significant data from large sets of information (Rodiyansyah, 2017). Hierarchical clustering can be characterized as an information gathering strategy that begins by clustering two items with the closest proximity or similarity in information. Meanwhile, non-hierarchical clustering is a strategy that starts by defining the ideal number of data groups in advance. Clustering is also known as a cycle in which information is gathered and partitioned into several datasets, allowing similar examples to form, which contain comparative information and are grouped together (Indraputra & Fitriana, 2020). The comparison between the two methods lies in the selection of the number of clusters. Non-hierarchical methods such as K-Means and Fuzzy C-Means (FCM) aim to partition data into one or more clusters, grouping data with similar characteristics together. The K-Means algorithm is generally very effective in partitional clustering and is one of the most commonly used methods, compared to other clustering methods due to its simplicity and efficiency. K-Means is a partitional algorithm, as it starts by specifying the initial number of clusters and defining the centroid value upfront (Syam, 2017). Research by Sholikhah (2022) compared clustering methods like K-Means, K-Medoids, X-Means, and DBSCAN, finding that K-Means was the best method for clustering districts. The study compares three clustering methods (K-Means, K-Medoids, and Random Clustering) to group elementary schools in Bojonegoro District based on educational capacity and facilities, finding that the K-Means method with five clusters provides the best solution (Nurdiansyah et al., 2023). The research of comparing the K-Means and K-Medoids clustering methods to group sub-districts in Bojonegoro Regency based on various types of Social Assistance, finding that the K-Means method with five optimal clusters provides the best results, with differences between 2020 and 2021 due to varying focuses on COVID-19 and direct regional and village assistance (Fitriyah et al., 2023). This research applies various K-Means clustering approaches, including Kernel K-Means and Fast K-Means, to group sub-districts in Bojonegoro District based on population data, finding that the Kernel K-Means method with five clusters provides the best results (Nurdiansyah et al., 2024).

Furthermore, clustering regions or districts based on tuberculosis cases and healthcare workers is crucial. Clustering helps the government identify areas that need more attention in reducing tuberculosis cases. According to previous studies, both K-Means and Fuzzy C-Means can provide the best clustering results. Both methods yield significant and efficient outcomes with different patterns. Therefore, this research will compare K-Means and Fuzzy C-Means to obtain the best accuracy results. Previous research compared the K-Means and Fuzzy C-Means (FCM) algorithms. For example, a study on grouping individuals with infectious diseases in Tuban Regency used K-Means, where infectious disease information was clustered into three, two, and four groups to identify similar and matching groups (Jaya & Oktafianto, 2022). Another study on clustering districts in Banyumas Regency using KCM and FCM showed that the groups' participation rates were similar. The clustering was based on population, healthcare workers, and health facilities. Scenario two showed better results than scenario one, with a variation ratio of 85% between clusters compared to the total variation (Jajang et al., 2021).

Fuzzy C-Means (FCM) has undergone significant advancements in recent years to enhance its accuracy and efficiency. For example, the integration of the Pareto principle in convergence strategies has reduced iteration counts by up to 75%, making it highly suitable for large-scale datasets like those in Big Data applications (Pérez-Ortega et al., 2024). In medical imaging, the adoption of Mahalanobis distance in FCM (FCM-M) has improved tumor segmentation in mammograms, supporting more accurate and early breast cancer detection (Krasnov et al., 2023). Furthermore, the addition of residual regulation to the objective function has made FCM more robust against noise, enabling its effective application in complex real-world image data (Wang et al., 2021).

This study will be conducted to answer how the descriptive statistics of tuberculosis cases and healthcare workers in Bojonegoro Regency are, as well as to compare two methods, K-Means and Fuzzy C-Means, for clustering sub-districts in Bojonegoro Regency based on tuberculosis cases and healthcare workers. Fuzzy C-Means is a calculation method that allows each individual to have varying degrees of participation in multiple groups (Rouza & Fimawahib, 2020). Fuzzy C-Means is also referred to as a clustering technique, where data membership is not solely determined by a single level of registration (Sormin et al., 2015). This method aims to limit the capability of the group and obtain relevant information within that group. On the other hand, K-Means is a technique for clustering objects into K groups or clusters. To perform this clustering, the value of K must be determined beforehand. Users generally have initial information about the objects being studied, including the appropriate number of clusters, and can use dissimilarity measures to define the objects (Selviana, 2016). In this study, programming with the R language was used. R software is a statistical software that is license-free and easy to get on the official website (Nurdiansyah & Sulistiawan, 2023).

In this study, using real data on the number of tuberculosis cases and healthcare workers in Bojonegoro Regency, a comparison between K-Means and Fuzzy C-Means methods will be made to cluster districts based on these characteristics. Dividing data into three different categories will help the Bojonegoro government in clustering districts with varying numbers of tuberculosis patients and healthcare workers. Additionally, the government can evaluate districts with high tuberculosis case numbers. Therefore, this research will be titled “Comparison of K-Means and Fuzzy C-Means for Optimizing Tuberculosis Management and Healthcare Service Allocation in Bojonegoro”.

METHOD

Research Design

This study employs a quantitative research design with an exploratory approach, aiming to cluster districts in Bojonegoro Regency based on the number of tuberculosis cases and the availability of healthcare workers. The clustering methods used are K-Means and Fuzzy C-Mean.

Population and Sample

The population in this study consists of tuberculosis cases and healthcare workers in Bojonegoro Regency. The research sample focuses on data regarding the number of tuberculosis cases and healthcare workers in all districts of Bojonegoro Regency in 2022.

Sampling Techniques

This quantitative study uses the purposive sampling technique, where data is collected based on the availability of information from publications by the Bojonegoro Regency Central Bureau of Statistics.

Research Subjects

The subjects of this study are tuberculosis cases and the number of medical personnel. The data used is secondary data sourced from publications by the Bojonegoro Regency Central Bureau of Statistics, consisting of data on tuberculosis cases and healthcare personnel across all districts in Bojonegoro Regency in 2022.

The research variables include the number of districts in Bojonegoro Regency, such as Ngraho, Tambakrejo, Ngambon, Ngasem, Bubulan, Dander, Sugihwaras, Kedungadem, Kepohbaru, Baureno, Kanor, Sumberrejo, Balen, Kapas, Bojonegoro, Kalitidu, Malo, Purwosari, Padangan, Kasiman, Temayang, Margomulyo, Trucuk, Sukosewu, Kedewan, Gondang, Sekar, and Gayam. Other variables include the number of healthcare personnel, consisting of medical staff, nursing staff, midwives, and pharmaceutical staff, as well as the number of tuberculosis cases in each district in Bojonegoro Regency.

Data Analysis Techniques

The analytical steps in this study were conducted to compare the K-Means and Fuzzy C-Means algorithms in clustering data, aiming to understand the characteristics of groups formed by each algorithm. This analysis seeks to determine the best clustering method for Bojonegoro Regency using data on tuberculosis cases and the number of healthcare workers in each district. The clustering steps used in this study are as follows (Giordani et al., 2020):

1. Import Data

   Import the data to be analyzed into R software. Ensure the data is in CSV (comma-delimited) format, which is compatible with MS Excel.
2. Determine Data Characteristics

   Identify the characteristics of tuberculosis cases and healthcare workforce data using descriptive statistics.
3. Data Preprocessing

   Perform preprocessing on the data, including:
   3.1. Removing missing values;
   3.2. Applying scaling or normalization;
   3.3. Converting data types as needed;
   3.4. Removing irrelevant columns.
4. Apply K-Means and Fuzzy C-Means Methods as discussed in Martos et al. (2023).
   4.1. The K-Means method is implemented using the **kmean()** function in R.
   4.2. The Fuzzy C-Means method is implemented using the **fcmean()** function.

   Parameters to configure include the number of clusters and the dataset to be used. The steps to implement these methods are as follows:
   a) Load the required library packages (**ClusterR** and **cluster**).
   b) Input the data.
   c) Run the **kmean()** or **fcmean()** program to determine clustering results.
   d) Display the clustering results.
5. Model Evaluation

   Evaluate the K-Means and Fuzzy C-Means models using SSE (Sum of Square Error).
6. Visualize Evaluation Results

   Plot an SSE line diagram for different K values for both K-Means and Fuzzy C-Means. The optimal K is determined using the Elbow Criterion, identifying the first significant "elbow" in the graph (Fitriyah et al., 2023).
7. Cluster Data

   Group the data based on the best clustering method determined through evaluation.
8. Prepare Data for Clustering

   Organize tuberculosis cases and healthcare workforce data, labeling it with clustering results and referencing the variables used for clustering.
9. Identify Cluster Characteristics

   Determine the characteristics of tuberculosis cases and healthcare workforce data based on the clustering results using descriptive statistics.

The analysis steps described above can be illustrated further using a flowchart following the research framework outlined in the reference (Pamungkas et al., 2021).

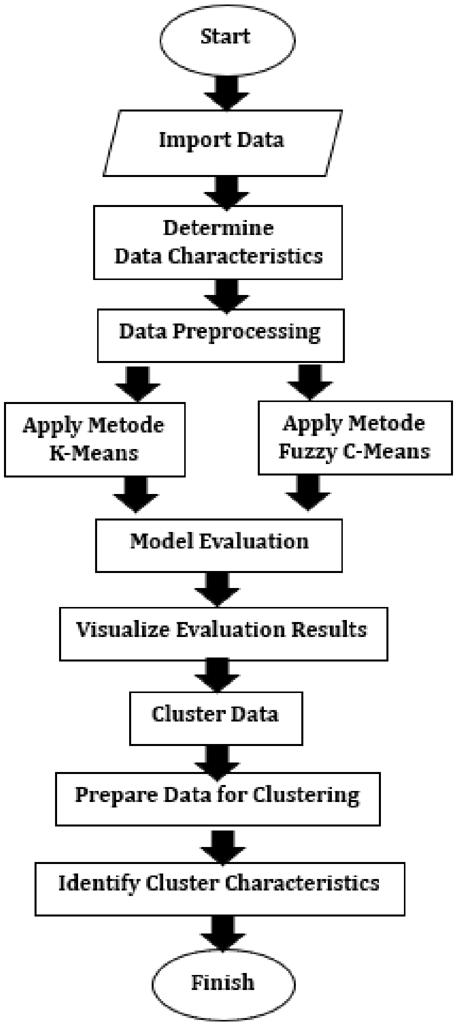

Figure 1. Flowchart Comparing the K-Means and Fuzzy C-Means Methods.

As a limitation regarding the selection of the best method in this study, the Sum of Square Error (SSE) measure was used, following the Elbow Criterion to determine the optimal number of clusters.

RESULTS AND DISCUSSION

Descriptive Statistics

In this study, descriptive statistics were analyzed using SPSS version 22. The results will be presented in Table 1 below.

Table 1. Descriptive Statistics of the Entire Dataset

Descriptive Statistics						
	N	Minimum	Maximum	Sum	Mean	Std. Deviation
Medical Personnel	28	1	11	110	3.93	2.276
Nursing Staff	28	7	34	424	15.14	6.900
Midwifery Staff	28	10	43	708	25.29	8.709
Pharmaceutical Staff	28	0	5	41	1.46	1.170
Tuberculosis Cases	28	5	152	1765	63.04	38.461

Table 1 presents a summary of the descriptive statistics for the availability of healthcare personnel, including medical staff, nursing staff, midwifery staff, pharmaceutical staff, and the number of tuberculosis cases in 28 districts of Bojonegoro Regency.

The lowest value for medical staff is 1, found in Trucuk District, while the highest value is 11, found in Kalitidu District. The total number of medical staff in Bojonegoro Regency is 105, with a mean of 3.93 and a standard deviation of 2.276.

For nursing staff, the lowest value is 7, found in Trucuk District, and the highest value is 34, found in Kedungadem District. The total number of nursing staff in Bojonegoro Regency is 423, with a mean of 15.14 and a standard deviation of 6.900.

The third variable, midwifery staff, has a lowest value of 10, found in Trucuk District, and the highest value of 43, found in Kedungadem District. The total number of midwifery staff in Bojonegoro Regency is 698, with a mean of 25.29 and a standard deviation of 8.709.

Next, for pharmaceutical staff, the lowest value is 0, found in Trucuk District, and the highest value is 5, found in Kalitidu District. The total number of pharmaceutical staff in Bojonegoro Regency is 41, with a mean of 1.46 and a standard deviation of 1.170.

The final variable is tuberculosis cases. The lowest number of tuberculosis cases is 5, found in Ngambon District, while the highest number is 152, found in Bojonegoro District. The total number of tuberculosis cases in Bojonegoro Regency is 1,765, with a mean of

63.04 and a standard deviation of 38.461.

Table 1 gives a clear picture of the availability of healthcare personnel and tuberculosis (TB) cases across 28 districts in Bojonegoro Regency. There are noticeable differences in the distribution of medical, nursing, midwifery, and pharmaceutical staff, as well as TB cases. For example, the number of medical staff ranges from just 1 in some districts to 11 in others, with an average of 3.93. TB cases show even greater variation, from as few as 5 cases to as many as 152, with an average of 63.04. These differences reveal inequalities in resource allocation, with districts like Trucuk consistently showing the lowest numbers, indicating a lack of adequate resources.

These variations play a vital role in clustering analysis, as districts with similar conditions naturally group together. For instance, districts with high TB cases may form a cluster that highlights areas with a significant disease burden, while those with fewer healthcare personnel may cluster as resource-scarce regions. This clustering helps policymakers identify patterns of inequality and design interventions that address the specific needs of each group effectively.

Comparison of Clustering Methods

The initial step before testing the dataset is to form clusters using the K-Means and Fuzzy C-Means methods to determine the Sum of Square Error (SSE). Graphs are created for both methods to identify and determine the optimal number of clusters, which can be observed as the first graph flattens out. The process of creating these graphs is presented in Figure 2 below:

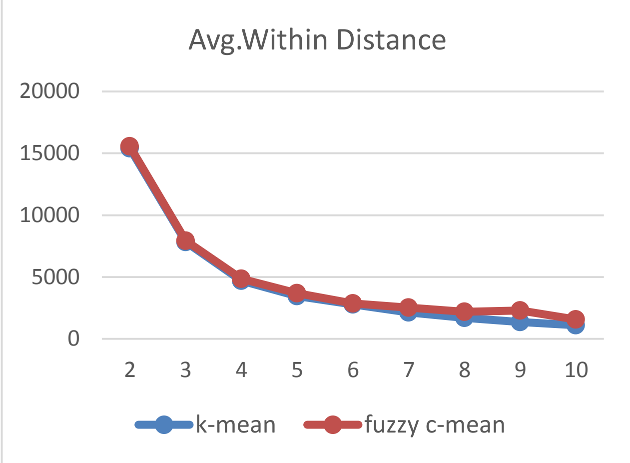

Figure 2. Comparison Graph of K-Means and Fuzzy C-Means Methods

From Figure 2, we can observe the comparison graph of the K-Means and Fuzzy C-Means methods based on healthcare personnel and tuberculosis case data in Bojonegoro Regency. The dataset, consisting of 28 data points, results in a cluster with a maximum of 5 clusters. It can be seen that Cluster 2 is at the highest point of the graph, indicating that using only two clusters is inefficient and does not adequately represent the data. Similarly, with three clusters, the error distance between clusters is still very high, meaning that using three clusters is not optimal for this analysis. However, at four clusters, the graph line starts to flatten and stabilize, suggesting that four clusters is the optimal point. Therefore, this number of clusters will be used in the comparison of the K-Means and Fuzzy C-Means methods for clustering districts in Bojonegoro Regency based on healthcare personnel and tuberculosis case data.

The comparison of clustering methods is performed by comparing the best average within distance value produced by each method. Based on the evaluation results from both methods, the comparison results can be seen in Table 2 below:

Table 2. Comparison Values of K-Means and Fuzzy C-Means Methods

	Avg.Within Distance	
k	**K-Mean**	**Fuzzy C-Mean**
2	15393.75	15563.02
3	7835.812	7930.687
4	**4704.031**	**4854.247**
5	3462.04	3702.436
6	2774.167	2867.369
7	2136.5	2534.702
8	1704.833	2189.86
9	1368.667	2286.91
10	1095.25	1574.649

From Table 3, it can be seen that with 4 clusters, the average within distance value is 4704.031 for the K-Means method and 4854.247 for the Fuzzy C-Means method. This indicates that the average within distance value for K-Means is lower than that of Fuzzy C-Means. Therefore, it can be concluded that in this study, the K-Means method performs better than the Fuzzy C-Means method.

Clustering Results of Districts in Bojonegoro Regency

Next, the results of the cluster search can be presented in an array, which is used to label each data point within the cluster. The array results from the data on the availability of healthcare personnel in all districts of Bojonegoro Regency are presented in Table 3 below:

Table 3. Clustering of Districts using the K-Means Method

Cluster	Sub-Districts
1	The sub-districts of Ngraho, Temayang, Sugihwaras, Sukosewu, Trucuk, Gayam, Malo, Purwosari, and Kasiman.
2	The sub-districts of Baureno, Bojonegoro, and Dander.
3	The sub-districts of Margomulyo, Ngambon, Sekar, Bubulan, Gondang, and Kedewan.
4	The sub-districts of Sumberrejo, Balen, and Kapas.
5	The sub-districts of Tambakrejo, Kedung Adem, Kepohbaru, Kanor, Ngasem, Kalitidu, and Padangan.

Based on Table 3, the data above shows that 28 districts are grouped into 5 main clusters. Cluster 1 consists of 9 districts: Ngraho, Temayang, Sugihwaras, Sukosewu, Trucuk, Gayam, Malo, Purwosari, and Kasiman. Cluster 2 includes only 3 districts: Baureno, Bojonegoro, and Dander. Cluster 3 comprises 6 districts: Margomulyo, Ngambon, Sekar, Bubulan, Gondang, and Kedewan. Cluster 4 contains Sumberrejo, Balen, and Kapas. Lastly, Cluster 5 consists of Tambakrejo, Kedungadem, Kepohbaru, Kanor, Ngasem, Kalitidu, and Padangan.

The next step is to identify the characteristics of each cluster, which are presented in Table 4 below:

Table 4. Characteristics of Each Cluster

Cluster	Sub-Districts	Characteristics
1	The sub-districts of Ngraho, Temayang, Sugihwaras, Sukosewu, Trucuk, Gayam, Malo, Purwosari, and Kasiman.	Cluster 1 has the highest number of medical personnel, nurses, midwives, and tuberculosis cases compared to other clusters.
2	The sub-districts of Baureno, Bojonegoro, and Dander.	Cluster 2 has a higher number of nurses and tuberculosis cases compared to Cluster 3 and Cluster 4.
3	The sub-districts of Margomulyo, Ngambon, Sekar, Bubulan, Gondang, and Kedewan.	Cluster 3 has the lowest number of medical personnel and tuberculosis cases compared to other clusters.
4	The sub-districts of	Cluster 4 has the

Cluster	Sub-Districts	Characteristics
	Sumberrejo, Balen, and Kapas.	lowest number of pharmaceutical personnel compared to other clusters.

Cluster 1, with the highest number of healthcare personnel and tuberculosis (TB) cases, represents areas with high service demand and a significant disease burden. Policies should focus on strengthening TB-specific interventions, such as increasing access to diagnostic tools, ensuring consistent medication supply, and conducting public health campaigns to reduce transmission. Additionally, training programs for healthcare personnel in effective TB management should be prioritized to improve patient care and outcomes.

Clusters 2, 3, and 4 highlight specific challenges requiring tailored solutions. Cluster 2, with moderate healthcare personnel but high TB cases, needs enhanced community-based TB prevention efforts, such as active case-finding and improved access to medical personnel. Cluster 3, with the lowest number of healthcare personnel and TB cases, should focus on maintaining preventive measures while improving baseline healthcare access through equitable staff redistribution. Lastly, Cluster 4, with the lowest number of pharmaceutical personnel, requires urgent efforts to improve medication availability and strengthen the pharmaceutical supply chain to ensure treatment accessibility, particularly for TB patients.

CONCLUSION

Conclusion

This study reveals variations in the distribution of healthcare workers and tuberculosis cases across 28 districts in Bojonegoro Regency. Clustering results show that the K-Means method outperforms Fuzzy C-Means, producing four distinct clusters. Cluster 1 has the highest number of healthcare workers and tuberculosis cases, while Cluster 3 has the lowest. These findings highlight the need for targeted attention to healthcare distribution and tuberculosis case management in each cluster.

This study highlights significant disparities in healthcare worker distribution and tuberculosis (TB) cases across Bojonegoro Regency. The clustering results provide a clear basis for targeted public health policies, with Cluster 1 requiring enhanced TB management resources and Cluster 3 needing improved access to basic healthcare services. These findings emphasize the importance of data-driven strategies to ensure equitable healthcare and effective TB control in each cluster.

Recommendations

Local governments are encouraged to improve the equitable distribution of healthcare workers, especially in underserved areas like Cluster 3, and prioritize tuberculosis management in high-case areas such as Cluster 1. Furthermore, the clustering results can be utilized for more efficient resource allocation and inspire further research by incorporating additional variables to enhance the analysis.

For future research, it is recommended to integrate additional variables such as socio-economic factors, environmental conditions, or transportation accessibility, as these may influence healthcare distribution and tuberculosis (TB) prevalence. Additionally, exploring hybrid clustering methods that combine the strengths of different algorithms could improve clustering accuracy and provide more robust insights into hidden patterns in healthcare data.

DAFTAR PUSTAKA

Anggraini, A. A., & Muharom, L. A. (2017). Pengelompokan Kecamatan Menggunakan Metode K-Means Cluster. Universitas Muhammadiyah Jember. http://repository.unmuhjember.ac.id/591/1/ARTIKEL.pdf

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

Giordani, P., Ferraro, M. B., & Martella, F. (2020). Fuzzy Clustering (pp. 125–211). https://doi.org/10.1007/978-981-13-0553-5_5

Hidayat, R., Wasono, R., & Darsyah, M. Y. (2017). Pengelompokan Kabupaten/Kota Di Jawa Tengah Menggunakan Metode K-Means Dan Fuzzy C-Means. Prosiding Seminar Nasional & Internasional, 240–250. https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/3017

Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. Jurnal Teknik Industri, 10(3), 275–282. https://doi.org/10.25105/jti.v10i3.8428

Jajang, J., Nurhayati, N., & Apriliana, Y. (2021). Aplikasi K-Means Dan Fuzy Clustering Dalam Pengelompokan Kecamatan Di Kabupaten Banyumas. Jurnal Ilmiah Matematika Dan Pendidikan Matematika, 13(2), 113. https://doi.org/10.20884/1.jmp.2021.13.2.5051

Jaya, K. I., & Oktafianto, K. (2022). Berdasrkan Kecamatan Di Kabupaten Tuban Menggunakan Algoritma K-Means. Curtina, 7(1), 22–31. https://doi.org/https://doi.org/10.55719/curtina.v3i1.441

Krasnov, D., Davis, D., Malott, K., Chen, Y., Shi, X., & Wong, A. (2023). Fuzzy C-Means Clustering: A Review of Applications in Breast Cancer Detection. Entropy, 25(7), 1021. https://doi.org/10.3390/e25071021

Mar'iyah, K., & Zulkarnain. (2021). Patofisiologi Penyakit Infeksi Tuberkulosis. In Prosiding Seminar Nasional Biologi, 7(1), 88–92. https://doi.org/https://doi.org/10.24252/psb.v7i1.23169

Martos, L. A. P., García-Vico, Á. M., González, P., & Carmona, C. J. (2023). Clustering: an R library to facilitate the analysis and comparison of cluster algorithms. Progress in Artificial Intelligence, 12(1), 33–44. https://doi.org/10.1007/s13748-022-00294-2

Mathofani, P. E., & Febriyanti, R. (2020). Faktor-Faktor Yang Berhubungan Dengan Kejadian Penyakit Tuberkulosis (TB) Paru di Wilayah Kerja Puskesmas Serang Kota Tahun 2019. JURNAL ILMIAH KESEHATAN MASYARAKAT : Media Komunikasi Komunitas Kesehatan Masyarakat, 12(1), 1–10. https://doi.org/10.52022/jikm.v12i1.53

Nurdiansyah, D., Ma'ady, M. N. P., Sukmawaty, Y., Utomo, M. C. C., & Mutiani, T. (2024). CLUSTERING ANALYSIS FOR GROUPING SUB-DISTRICTS IN BOJONEGORO DISTRICT WITH THE K-MEANS METHOD WITH A VARIETY OF

APPROACHES. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 18(2), 1095–1104. https://doi.org/10.30598/barekengvol18iss2pp1095-1104

Nurdiansyah, D., Saidah, S., & Cahyani, N. (2023). DATA MINING STUDY FOR GROUPING ELEMENTARY SCHOOLS IN BOJONEGORO REGENCY BASED ON CAPACITY AND EDUCATIONAL FACILITIES. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 17(2), 1081–1092. https://doi.org/10.30598/barekengvol17iss2pp1081-1092

Nurdiansyah, D., & Sulistiawan, A. (2023). Dasar Pemrograman Komputer Dengan Open Source Software R (Untuk Bidang Sains dan Teknologi). CV. AA. RIZKY.

Pamungkas, M. A., Oktavianto, H., & Umilasari, R. (2021). Perbandingan Fuzzy C-Means Dan K-Means Untuk Mengelompokkan Tingkat Buta Huruf Berdasarkan Provinsi Di Indonesia. http://repository.unmuhjember.ac.id/8837/

Pérez-Ortega, J., Moreno-Calderón, C. F., Roblero-Aguilar, S. S., Almanza-Ortega, N. N., Frausto-Solís, J., Pazos-Rangel, R., & Rodríguez-Lelis, J. M. (2024). A New Criterion for Improving Convergence of Fuzzy C-Means Clustering. Axioms, 13(1), 35. https://doi.org/10.3390/axioms13010035

Ramadhani, L., Purnamasari, I., & Amijaya, F. D. T. (2018). Penerapan Metode Complete Linkage dan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016). Jurnal EKSPONENSIAL, 9(1), 1–9. https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/208

Ramadhani, R. D. (2014). Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. Industrial Marketing Management, 1(1), 1–9. https://core.ac.uk/download/pdf/35378856.pdf

Rodiyansyah, S. F. (2017). K-Means Dan Fuzzy C-Means Pada Analisis Data Polusi Udara Di Kota X. Seminar Nasional Teknologi Informasi Dan Multimedia 2017, 25–30. https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1689

Rouza, E., & Fimawahib, L. (2020). Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu. Techno.Com, 19(4), 481–495. https://doi.org/10.33633/tc.v19i4.4101

Sejati, A., & Sofiana, L. (2015). Faktor-Faktor Terjadinya Tuberkulosis. Jurnal Kesehatan Masyarakat, 10(2), 122. https://doi.org/10.15294/kemas.v10i2.3372

Selviana, N. I. (2016). Analisis perbandingan K-Means dan Fuzzy C-Means untuk pemetaan motivasi balajar mahasiswa. Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 8, 01(01), 95–105. https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/2815

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

Sormin, R. P. A., Rumlawang, F. Y., & Sinay, L.

J. (2015). Aplikasi Metode Fuzzy C-Means Untuk Pengklasteran Kelayakan Rumah Di Desa Wayame, Ambon. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 9(2), 135–146. https://doi.org/10.30598/barekengvol9iss2pp135-146

Syam, F. A. (2017). Implementasi Metode Klastering K-Means untuk Mengelompokan Hasil Evaluasi Mahasiswa. Jurnal Ilmu Komputer Dan Bisnis, 8(1), 1857–1864. https://doi.org/10.47927/jikb.v8i1.94

Wang, C., Pedrycz, W., Li, Z., & Zhou, M. (2021). Residual-driven Fuzzy C-Means Clustering for Image Segmentation. IEEE/CAA Journal of Automatica Sinica, 8(4), 876–889. https://doi.org/10.1109/JAS.2020.1003420