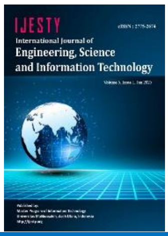
International Journal of Engineering, Science and InformationTechnology
Volume 5, No. 1 (2025) pp. 177-184
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i1.679
Research Paper, Short Communication, Review, Technical Paper



Application of K-Medoids Clustering Method on Disease Clustering Based on Patient Medical Records

Dian Fatika*, Bustami, Yesy Afrillia

Department of Informatics, Universitas Malikussaleh, Aceh, Indonesia

**Corresponding author Email: dian.200170231@mhs.unimal.ac.id*

The manuscript was received on 25 June 2024, revised on 28 August 2024, and accepted on 15 December 2024, date of publication 9 January 2025

Abstract

Dr. Fauziah Bireuen Regional General Hospital (RSUD) faces daily challenges in managing the ever-increasing medical record data. Currently, the medical record data only consists of reports containing information on the number of patients and their diseases, which are then archived without further processing to generate valuable information. This research aims to cluster diseases based on patient medical records using the K-Medoids Clustering method, thereby providing information on the patterns of disease spread in various regions of the Bireuen Regency. The data used are patient medical records from RSUD Dr. Fauziah Bireuen from 2021–2023, focusing on five common diseases: stroke, hypertension, schizophrenia, dyspepsia, and pneumonia. We conducted Clustering in 17 sub-districts in Bireuen Regency using the K-Medoids method and determined the optimal number of clusters using the Elbow method. The research results show that the K-Medoids method successfully grouped each disease into 3 clusters: high, medium, and low. The results showed that the K-Medoids method successfully grouped each disease into 3 clusters: high, medium, and low. The cluster distribution for stroke disease consists of 7 sub-districts in the high cluster, 7 in the medium, and 3 in the low. Hypertension disease consists of 6 sub-districts in the high cluster, 3 in the medium, and 8 in the low. Schizophrenia disease comprises seven sub-districts in the high cluster, 8 in the medium, and 2 in the low. Dyspepsia disease includes six sub-districts in the high cluster, 2 in the medium, and 9 in the low. Meanwhile, pneumonia disease consists of 8 sub-districts in the high cluster, 5 in the medium, and 4 in the low.

Keywords: *Clustering, K-Medoids, Medical Record, Elbow.*

1. Introduction

Several factors, such as genetics, environment, quality of health care, and individual behaviour, can influence diseases in various ways. As global interconnection increases, the spread of diseases becomes faster and broader. Therefore, efforts to categorize diseases are strategic steps to detect, prevent, and control the spread of these diseases. Based on the 2018 Riskesdas data, diseases with the highest prevalence include malnutrition (11.4%), stroke (10.9%), hypertension (8.36%), joint disease (7.30%), and diarrhea (6.8%). This condition affects the affected individuals and the overall capacity of health services [1]. Several factors, such as genetics, environment, quality of health services, and individual behaviour, can influence diseases with various variations. As global interconnectedness increases, the spread of diseases becomes more rapid and widespread. Therefore, efforts to categorize diseases are strategic steps to detect, prevent, and control the spread of these diseases. Based on the 2018 Riskesdas data, diseases with the highest prevalence include malnutrition (11.4%), stroke (10.9%), hypertension (8.36%), joint disease (7.30%), and diarrhea (6.8%). This condition impacts the affected individuals and affects the overall capacity of health services [1].

Dr. Fauziah Hospital in Bireuen Regency, as a central health facility in the region, faces the challenge of managing the growing amount of medical record data. Currently, medical record data at the hospital is only presented as a report of the number of patients and types of diseases without further processing to generate strategic insights, such as common disease patterns in the community [2]. This limitation hinders policy-making that is based on in-depth data. With the increasing volume of data, manual management is becoming increasingly ineffective and requires a technology-based approach for more structured data analysis [3].

As a solution, this research uses the K-Medoids Clustering method, one of the techniques in Data Mining, to cluster medical record data based on patterns and similarities in disease types. This method was chosen because of its ability to handle heterogeneous data and produce accurate clustering [4]. The results of this method of analysis are expected to provide helpful information about the level of disease risk in various regions, support hospital policies, and improve efficiency in prevention, treatment, drug stock management, and medical facility planning [5].
This research also contributes to previous literature that applies the K-Medoids method to health data. For example, Pohan et al. (2021) successfully categorized the prevalence of stunting in toddlers in Indonesia, Mustikawati et al. (2023) mapped family health indicators in Bondowoso Regency, and Munandar (2022) clustered poverty levels in Banten Province [6][7]. These studies show that the K-Medoids method effectively analyses complex data and provides reliable clustering results [8]. Thus, applying this method to the medical record data of RSUD, Dr Fauziah is expected to provide relevant insights to support the digital transformation of health services and improve service quality through more in-depth data analysis.

2. Literature Review

In recent years, using the K-Medoids algorithm in clustering health data has shown promising results. Research by Minarni et al. (2021) applied the K-Medoids algorithm to cluster data on patients with dengue fever, tuberculosis, and pneumonia in Agam Regency, West Sumatra, using data from 2016 to 2019. The results of this study identified two clusters, namely low and high clusters, with details that Tuberculosis disease has 32 members in the high cluster and 60 members in the low cluster, Pneumonia disease with 33 members in the high cluster and 59 members in the low cluster, and DHF disease with 10 members in the high cluster and 82 members in the low cluster. These findings demonstrate the effectiveness of the K-Medoids algorithm in clustering disease data based on different prevalence [9].
In addition, research by Bella Nurseptia et al. (2023) explored the application of the K-Medoids algorithm to map the level of violence against children and women in DKI Jakarta. Using the Davies Bouldin Index (DBI) method for evaluation, this study found that clustering the areas resulted in three clusters based on the level of violence: one cluster with a low level of violence (8 sub-districts), one with a medium level (30 sub-districts), and one with a high level (6 sub-districts). This result confirms the K-Medoids algorithm's ability to understand better the distribution of violence cases based on geographical location [10].
K-Means method to cluster the medical record data of BPJS Health user patients. Despite the focus on the K-Means algorithm, this study provides insights into the pattern of disease spread based on the class of service with three clusters formed. Nonetheless, it is essential to note that K-medoids are often superior in dealing with data containing noise and outliers compared to K-means [11].
Overall, these studies show that the K-Medoids algorithm is an effective method for clustering health data, enabling the identification of disease patterns that can be used to formulate better health policies [12]. With the ability to handle outliers and provide more stable results than other methods, such as K-Means, K-Medoids are an attractive option in health data analysis. Further research is needed to explore the application of this method in different contexts and compare it with other clustering techniques in various health settings.

3. Research Methods

Data Mining is discovering hidden patterns or valuable information from large amounts of data. By utilizing techniques from various fields, such as machine learning, pattern recognition, statistics, databases, and data visualization, Data Mining aims to generate insights that support more effective decision-making [13]. Techniques such as artificial intelligence, statistical analysis, and machine learning are used to identify relevant patterns that can be applied in various fields, including market analysis and consumer behaviour prediction [14]. Data mining is part of the knowledge discovery in database (KDD) process, which includes stages such as data collection, transformation, and analysis. This framework allows for systematically extracting meaningful information from large data sets [15].
The Elbow method determines the optimal number of clusters in clustering analysis by identifying the elbow points on the graph. This technique seeks an efficient number of clusters, where adding clusters no longer significantly reduces the Sum of Square Error (SSE) value. The Elbow method graph aids visualization, with the elbow point indicating the optimal cluster that produces the lowest internal variance [16].

$$SSE = \sum_{i=1}^{k} \sum_{x_i \in C_i} D(x_i\, C_i)^2 \qquad \text{.....(1)}$$

Clustering is a method to group objects based on certain similarities to form clusters, which are groups with similar objects but different from objects in other clusters [17]. The goal is to divide the dataset into groups that are homogeneous within and heterogeneous between groups. In Data Mining, Clustering helps find data distribution patterns to support further analysis. Similarity between objects is usually calculated based on the proximity of attributes in a multidimensional space [18].
Clustering methods consist of Hierarchical Clustering and Non-Hierarchical Clustering. Hierarchical Clustering forms a multilevel structure resembling a tree by grouping the most similar objects in order until a complete hierarchy is formed. In contrast, Non-Hierarchical Clustering requires determining the number of clusters upfront and then grouping the data without regard to hierarchy [19].
The Partitioning Around Medoid (PAM) algorithm, also known as K-Medoids, was developed by Leonard Kaufman and Peter J. Rousseeuw in 1987. This algorithm uses medoids as cluster centres, which are the original objects in the dataset and represent the cluster. K-medoids are similar to K-means but have a key difference in determining the cluster centre. K-medoids use the original object as the cluster centre, while K-means use the average value (mean) of all objects in the cluster. As a result, cluster centres in K-Medoids are always accurate data, while in K-Means, they are calculated values [20]. The main advantage of K-Medoids is their ability to overcome the weaknesses of K-Means, such as sensitivity to outliers or extreme values that can affect clustering results. In addition, K-Medoids produce more stable results as they do not depend on the order of the input data.
The K-Medoids algorithm starts by initializing k cluster centres (the number of clusters) and allocating each object to the nearest cluster using the Euclidean Distance, which is calculated by the following formula:

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

The scheme of the disease clustering system based on patient medical records using the K-Medoids method is described as follows :

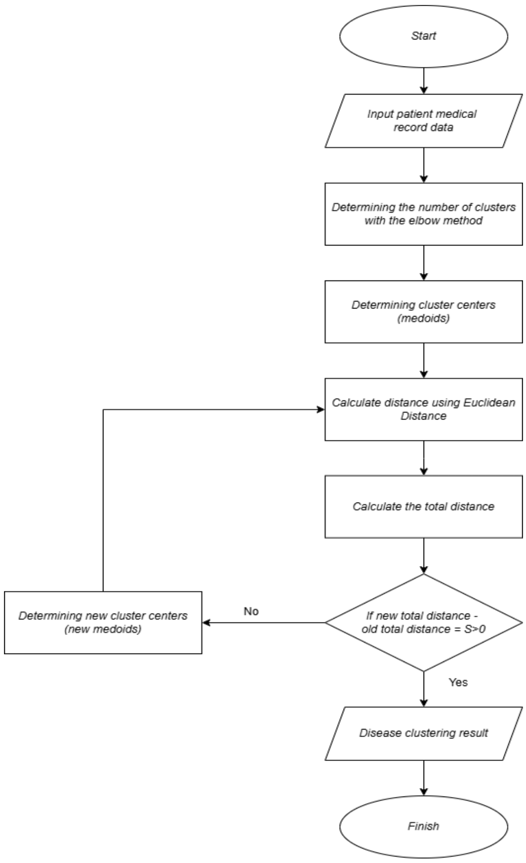


Fig 1. System Schematic

The system starts by entering the patient's medical record data and determining the number of clusters to be formed (k) using the Elbow method. The initial cluster centres of k are then randomly selected. Next, each data is allocated to the nearest cluster based on the Euclidean distance calculation. After that, the total distance is calculated, followed by the deviation calculation, which is the difference between the new and previous total distances. If the deviation (S) is less than zero, the process is repeated until the deviation is greater than zero. When (S>0), the clustering process is stopped, and the system generates clusters based on the patient's medical record. The final result is a grouping of diseases into three categories: high cluster, medium cluster, and low cluster.

4. Results and Discussion

This study analyzed patient medical record data at RSUD Dr. Fauziah Bireuen from 2021 to 2023. The data used included five types of diseases, namely stroke, hypertension, schizophrenia, dyspepsia, and pneumonia. The analysis was conducted in 17 sub-districts in Bireuen Regency, including Kota Juang, Jeumpa, Kuala, Juli, Jeunieb, Peulimbang, Pandrah, Peudada, Samalanga, Simpang Mamplam, Jangka, Peusangan, Peusangan Selatan, Peusangan Siblah Krueng, Makmur, Kutablang, and Gandapura. Once the data was collected, the elbow method was applied to determine the number of clusters, followed by the K-Medoids algorithm calculation process using the Euclidean Distance formula. The analysis results show different cluster patterns for each type of disease.

In stroke disease, the cluster results are C1, as many as seven sub-districts; C2, as many as seven sub-districts; and C3, as many as three sub-districts. For hypertension, the cluster results consist of C1 in 6 sub-districts, C2 as many as three sub-districts, and C3 as many as eight sub-districts. In schizophrenia, clustering results in C1 in seven sub-districts, C2 in eight sub-districts, and C3 in as many as two sub-districts. For dyspepsia, the cluster results are C1, as many as six sub-districts; C2, as many as two sub-districts; and C3, as many as nine sub-districts. Finally, for pneumonia, the cluster results show C1 in as many as eight subdistricts, C2 in as many as five subdistricts, and C3 in as many as four subdistricts. This study provides an overview of disease distribution patterns based on geographical location in the Bireuen District.

4.1. Cluster Analysis Results

The following are the results and analysis of all cluster results on the disease-type dataset:

4.1.1. Stroke Disease Cluster Results

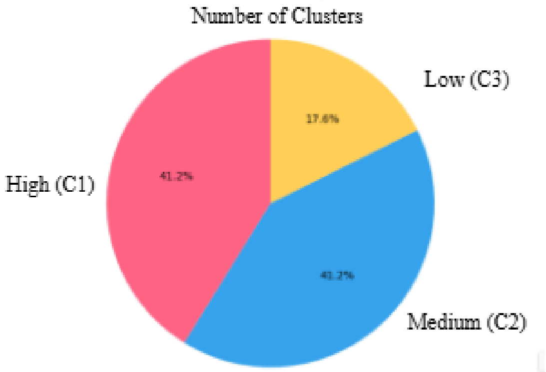


Fig 2. Stroke Disease Cluster Visualization

Table 1. Cluster Distribution of Stroke Disease

No	Cluster Type	Number of Sub-districts	Percentage (%)
1	High(C1)	7	41,2 %
2	Medium(C2)	7	41,2 %
3	Low (C3)	3	17,6 %

Based on data from the last three years, stroke cases in Bireuen District are divided into three clusters. The High Cluster (C1) includes seven sub-districts: Kota Juang, Jeumpa, Jeunieb, Simpang Mamplam, Peusangan, Makmur, and Gandapura, which require severe treatment as cases remain high or are increasing. Medium Cluster (C2) includes seven sub-districts: Kuala, Juli, Peulimbang, Peudada, Samalanga, Jangka, and Kuta Blang, which require monitoring to prevent further increase. Meanwhile, the Low Cluster (C3) consists of 3 sub-districts: Pandrah, Peusangan Selatan, and Peusangan Siblah Krueng, which need to focus on prevention and education to keep cases low.

4.1.2. Hypertension Disease Cluster Results

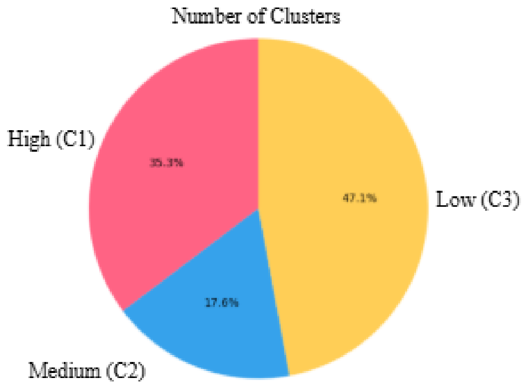


Fig 3. Hypertension Disease Cluster Visualization

Table 2 Cluster Distribution of Hypertension Disease

No	Cluster Type	Number of Sub-districts	Percentage (%)
1	High(C1)	6	35,3 %
2	Medium (C2)	3	17,6 %
3	Low (C3)	8	47,1 %

Based on hypertension distribution data, the High Cluster (C1) includes six sub-districts: Kota Juang, Jeumpa, Kuala, Juli, Jangka, and Peusangan, which require serious attention due to the increasing risk of cases. The Medium Cluster (C2) includes three sub-districts: Peudada, Simpang Mamplam, and Kuta Blang, which requires monitoring to prevent further increases. The Low Cluster (C3) consists of 8 sub-districts: Jeunieb, Peulimbang, Pandrah, Samalanga, Peusangan Selatan, Peusangan Siblah Krueng, Makmur, and Gandapura, which need to focus on prevention to keep prevalence low. Intensive prevention, monitoring and treatment strategies in each cluster can help reduce the district's hypertension burden.

4.1.3. Schizophrenia Disease Cluster Results

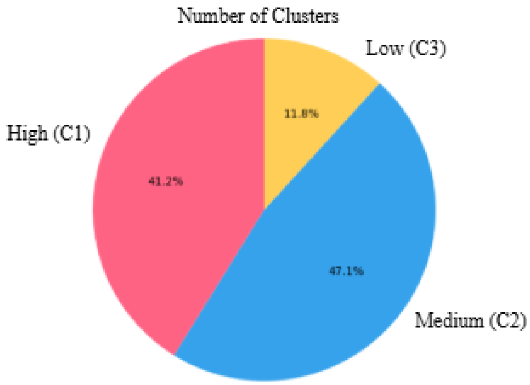


Fig 4. Schizophrenia Disease Cluster Visualization

Table 3. Cluster Distribution of Schizophrenia Disease

No	Cluster Type	Number of Sub-districts	Percentage (%)
1	High (C1)	7	41,2 %
2	Medium (C2)	8	47,1 %
3	Low (C3)	2	11,8 %

Based on schizophrenia distribution data, the High Cluster (C1) covers seven sub-districts: Kota Juang, Kuala, Juli, Peudada, Simpang Mamplam, Jangka, and Peusangan. These sub-districts require serious attention due to the high risk of schizophrenia, potentially leading to an increase in cases each year. The Medium Cluster (C2) consists of 8 sub-districts: Jeumpa, Jeunieb, Pandrah, Samalanga, Peusangan Siblah Krueng, Makmur, Kuta Blang, and Gandapura. Meanwhile, the Low Cluster (C3) includes 2 sub-districts, Peulimbang and Gandapura, which require a focus on prevention to keep prevalence low.

4.1.4. Dyspepsia Disease Cluster Results

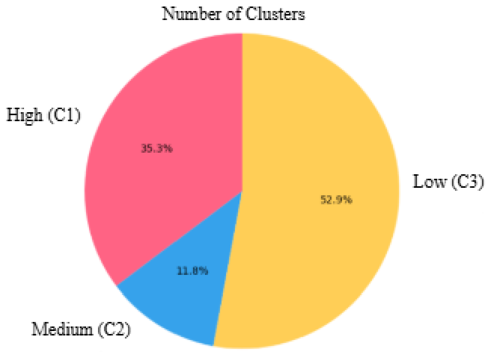


Fig 5. Dyspepsia Disease Cluster Visualization

Table 4. Cluster Distribution of Dyspepsia Disease

No	Cluster Type	Number of Sub-districts	Percentage (%)
1	High (C1)	7	35,3%
2	Medium (C2)	8	11,8%
3	Low (C3)	2	52,9%

Based on dyspepsia distribution data, the High Cluster (C1) covers six sub-districts: Kota Juang, Jeumpa, Kuala, Juli, Jangka, and Peusangan, which require serious attention due to the high risk of dyspepsia. The Medium Cluster (C2) consists of 2 sub-districts, Peudada and Samalanga, which require monitoring to prevent increased cases. Low Cluster (C3) includes nine sub-districts: Jeunieb, Peulimbang, Pandrah, Simpang Mamplam, South Peusangan, Peusangan Siblah Krueng, Makmur, Kuta Blang, and Gandapura, which need to focus on prevention to keep prevalence low. Although most sub-districts have low prevalence, clusters with high prevalence indicate the need for special attention and strategic measures to reduce the risk of dyspepsia in the region.

4.1.5. Pneumonia Disease Cluster Results

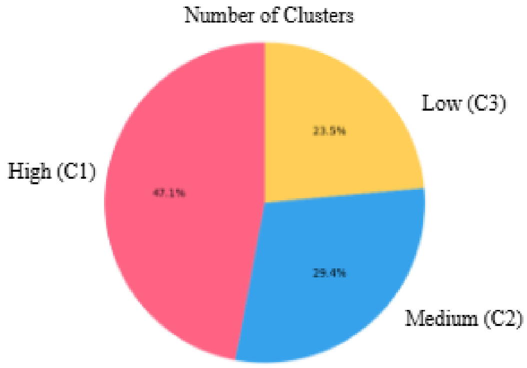


Fig. 6 Pneumonia Disease Cluster Visualization

Table 5. Cluster Distribution of Pneumonia Disease

No	Cluster Type	Number of Sub-districts	Percentage (%)
1	High (C1)	8	47,1 %
2	Medium (C2)	5	29,4 %
3	Low (C3)	4	23,5 %

Based on pneumonia distribution data, the High Cluster (C1) covers eight sub-districts: Kota Juang, Jeumpa, Kuala, Juli, Jeunieb, Peudada, Peusangan, and Gandapura, which require serious attention due to the high risk of disease. The Medium Cluster (C2) comprises five sub-districts: Peulimbang, Simpang Mamplam, South Peusangan, Peusangan Siblah Krueng, and Kuta Blang. Low Cluster (C3) includes four sub-districts: Pandrah, Samalanga, Jangka, and Makmur, which need to focus on prevention to keep prevalence low. The high spread of pneumonia in almost half of Bireuen district requires quick and strategic steps from health authorities, including prevention efforts, effective treatments, and improved health facilities.

4.2. System Implementation

Using Python technology with the Flask framework and MySQL database management, a web application has been developed that provides features to facilitate the analysis and monitoring of medical disease distribution areas. A summary of regions and data that displays the distribution of clusters in each region and a graph that visualizes the number of distributions of each cluster.

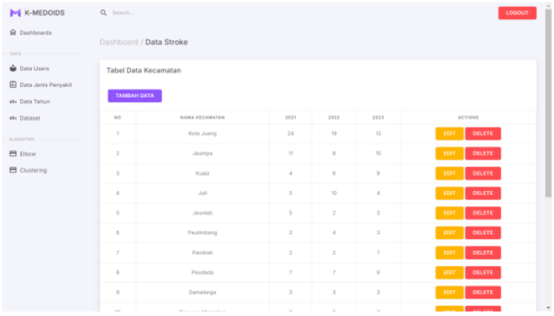


Fig 7. Dataset page

Fig 8. Elbow Method page

This Dataset and Elbow Method page is designed to display data from each sub-district in detail, with options to add, edit, or delete data as needed. In addition, this page presents the analysis results using the elbow method, which is used to determine the optimal number of clusters in grouping data, making it easier to understand the distribution patterns of diseases in each sub-district. These features provide flexibility and accuracy in managing and analyzing health data.

Fig 9. Clustering page

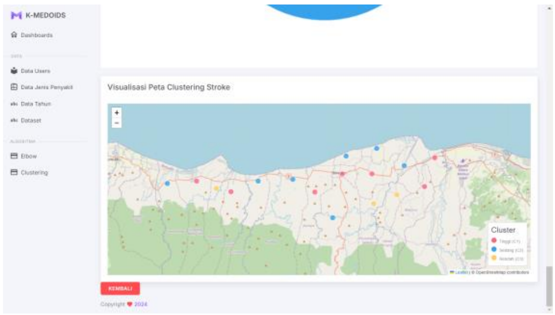
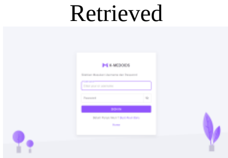
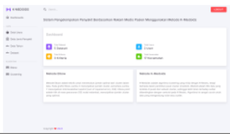
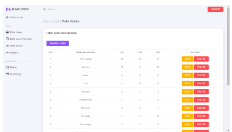


Fig 10. Clustering Mapping Visualization Page

This clustering and cluster mapping visualization page allows users to perform the clustering process and view the results directly. The clustering results are visualized through interactive graphs and maps, including bar charts, to make it easier for users to understand the distribution of clusters and the spread of disease in each region. This visualization is designed to provide a clear and intuitive picture, thus helping users to analyze data more effectively and support data-driven decision-making.

System testing executes software to ensure the system complies with the specified specifications and runs well in the expected environment. The results of this test will be displayed in tabular form to facilitate evaluation and analysis of system performance.

Table 6. System Testing

NO	Testing	Test Case	Expected Result	Test Result	Conclusion
1	Form Login	Input Username and password	Home View	Retrieved 	Valid
2	Page Home	Click Home	Display Information	Retrieved 	Valid
3	Page Dataset	Click Dataset	Show, add data	Retrieved 	Valid
4	Page Method Elbow	Click Page elbow	Displaying Elbow Method Results	Retrieved	Valid

					
5	Page Clustering	Click Page Clustering	Display Clustering results and visualizations	Retrieved 	Valid

The table above results from system testing to ensure each feature works according to specifications and user needs.

a. Login Form: When the user enters the username and password, the system successfully redirects to the Home page as expected, so this test is valid.
b. Home Page: When the "Home" button is clicked, the information is displayed correctly according to the expected function, so this test is valid.
c. Dataset page: Testing on the Dataset shows that the features for displaying and adding data work as expected, so it is declared valid.
d. Elbow Method Page: When the user accesses the Elbow method page, the method results are displayed correctly, making this test valid.
e. Clustering Page: This page successfully displays the expected clustering results and visualization, so this test is also valid.
f. The test results show that all system features run well and follow the designed specifications.

5. Conclusion

Several conclusions can be drawn from the research conducted on disease clustering based on patient medical records at Dr. Fauziah Hospital using the K-Medoids method. This study selected five types of diseases that were most commonly suffered based on medical record data from 2021-2023, namely stroke, hypertension, schizophrenia, dyspepsia, and pneumonia. Clustering was conducted based on 17 sub-districts in Bireuen Regency, namely Kota Juang, Jeumpa, Kuala, Juli, Jeunieb, Peulimbang, Pandrah, Peudada, Samalanga, Simpang Mamplam, Jangka, Peusangan, Peusangan Selatan, Peusangan Siblah Krueng, Makmur, Kutablang, and Gandapura, to determine the distribution of these diseases.
Applying the K-Medoids clustering method resulted in the three most optimal clusters according to the Elbow method: high, medium, and low. The distribution of stroke disease showed a high cluster (C1) of 41.2% with seven sub-districts, a medium cluster (C2) of 41.2% with seven sub-districts, and a low cluster (C3) of 17.6% with three sub-districts. The distribution of hypertension disease resulted in a high cluster (C1) of 31.3% with six sub-districts, a medium cluster (C2) of 17.6% with three sub-districts, and a low cluster (C3) of 47.1% with eight sub-districts. Schizophrenia disease distribution shows a high cluster (C1) of 41.2% with seven sub-districts, a medium cluster (C2) of 47.1% with eight sub-districts, and a low cluster (C3) of 11.8% with two sub-districts. The distribution of dyspepsia disease resulted in a high cluster (C1) of 35.3% with six sub-districts, a medium cluster (C2) of 11.8% with two sub-districts, and a low cluster (C3) of 52.9% with nine sub-districts. The distribution of pneumonia showed a high cluster (C1) of 47.1% with eight subdistricts, a medium cluster (C2) of 29.4% with five subdistricts, and a low cluster (C3) of 23.5% with four subdistricts.
Based on the clustering results, it is known that the disease with the highest percentage in the high cluster is pneumonia, which requires special attention so that the number of sufferers does not continue to grow. The disease with the highest percentage in the medium cluster is schizophrenia, which requires routine monitoring and improved health services. Meanwhile, the disease with the highest percentage in the low cluster is dyspepsia, which requires educational measures to ensure cases remain low.

References

[1] Kemenkes, "Laporan Riskesdas 2018 Nasional.pdf," Lembaga Penerbit Balitbangkes. p. hal 156, 2018.
[2] PERMENKES RI No 269/MENKES/PER/III/2008, "permenkes ri 269/MENKES/PER/III/2008," Permenkes Ri No 269/Menkes/Per/Iii/2008, vol. 2008. p. 7, 2008.
[3] Angelina M. T. I. Sambi Ua et al., "Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru," J. Publ. Tek. Inform., vol. 2, no. 2, pp. 88–99, 2023, doi: 10.55606/jupti.v2i2.1742.
[4] M. N. P. Pamulang, M. N. Aini, and U. Enri3, "Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokkan Penyakit ISPA," Edumatic J. Pendidik. Inform., vol. 5, no. 1, pp. 99–107, 2021, doi: 10.29408/edumatic.v5i1.3359.
[5] D. Permata Sari, "Pengelompokkan Penyakit Berdasarkan Lingkungan Dengan Algoritma K-Means Pada Puskesmas Sungai Tarab 2," JOISIE (Journal Inf. Syst. Informatics Eng., vol. 5, no. 2, pp. 75–81, 2021, doi: 10.35145/joisie.v5i2.1700.
[6] H. Pohan, M. Zarlis, E. Irawan, H. Okprana, and Y. Pranayama, "Penerapan Algoritma K-Medoids dalam Pengelompokan Balita Stunting di Indonesia," JUKI J. Komput. dan Inform., vol. 3, no. 2, pp. 97–104, 2021, doi: 10.53842/juki.v3i2.69.
[7] E. M. P. Hermanto, H. B. Rochmanto, and R. Agustin, "Pemetaan Program Indonesia Sehat dengan Pendekatan Keluarga (PIS PK) di Kabupaten Bondowoso dengan K-Medoids," J. Stat. dan Komputasi, vol. 2, no. 2, pp. 83–92, 2023, doi: 10.32665/statkom.v2i2.2307.
[8] T. A. Munandar, "Penerapan Algoritma Clustering Untuk Pengelompokan Tingkat Kemiskinan Provinsi Banten," JSiI (Jurnal Sist. Informasi), vol. 9, no. 2, pp. 109–114, 2022, doi: 10.30656/jsii.v9i2.5099.

[9] M. Minarni, E. I. Sari, A. Syahrani, and P. Mandarani, “Klasterisasi Penyakit Menggunakan Algoritma K-Medoids pada Dinas Kesehatan Kabupaten Agam,” J. Nas. Pendidik. Tek. Inform., vol. 10, no. 3, p. 137, 2021, doi: 10.23887/janapati.v10i3.34904.

[10] B. Nurseptia, N. Sulistiyowati, and A. Voutama, “Pemetaan Tingkat Kekerasan Pada Anak Dan Perempuan Menggunakan Algoritma K-Medoids (Studi Kasus: P2TP2A DKI Jakarta),” vol. 10, no. 4, pp. 244–255, 2023.

[11] J. Wandana, S. Defit, and S. Sumijan, “Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means,” J. Inf. dan Teknol., vol. 2, pp. 4–9, 2020, doi: 10.37034/jidt.v2i4.73.

[12] A. Suprianto, H. Latipa Sari, and R. Zulfiandry, “Perbandingan Algoritma K-Means Dan K-Medoid Dalam Pengelompokan Data Pasien Berdasarkan Rekam Medis Di Puskesmas M. Thaha Bengkulu Selatan,” J. Sci. Soc. Res., vol. 4307, no. 3, pp. 580–586, 2023.

[13] R. T. S. Muhammad Hariyanto, “Clustering pada Data Mining untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance,” Sist. Komput. dan Tek. Inform., vol. 1, no. 1, pp. 117–122, 2018.

[14] L. Rosnita, Y. Afrillia, R. P. Fhonna, and U. Ilyatin, "Development of Web-Based Tracer Alumni Information System," J. Comput. Sci. Inf. Technol. Telecommun. Eng., vol. 2, no. 2, pp. 202–210, 2021, doi: 10.30596/jcositte.v2i2.7845.

[15] R. Bayu Prasetyo, Y. Agus Pranoto, and R. Primaswara Prasetya, “Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Penyakit Pasien Rawat Jalan Pada Klinik Dr. Atirah Desa Sioyong, Sulteng,” JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 4, pp. 2144–2151, 2023, doi: 10.36040/jati.v7i4.7419.

[16] N. A. Maori and E. Evanita, “Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering,” Simetris J. Tek. Mesin, Elektro dan Ilmu Komput., vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

[17] S. Sidabutar, Edpid. 2020.

[18] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” J. Teknol. dan Sist. Inf., vol. 2, no. 2, p. 100, 2021.

[19] N. Nurdin, S. Fitriani, Z. Yunizar, and B. Bustami, "Clustering the Distribution of COVID-19 in Aceh Province Using the Fuzzy C-Means Algorithm," JTAM (Jurnal Teor. dan Apl. Mat., vol. 6, no. 3, p. 665, 2022, doi: 10.31764/jtam.v6i3.8576.

[20] N. R. Aeni, A. Nilogiri, and R. Umilasari, “Algoritma Partitioning Around Medoids Dalam Mengelompokkan Provinsi Di Indonesia Berdasarkan Indeks Kinerja Davies Bouldin Pada Kasus Penyakit HIV,” pp. 1–9, 2020.