JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 733 – 740 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v11i4.4140
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

COMPARISON OF K-MEANS AND K-MEDOIDS FOR DRUG DATA CLUSTERING

Tripa Andika[1*], Kurniabudi [1], Sharipuddin[1]
[1]Information Systems, Universitas Dinamika Bangsa
email: *andikatrifa@gmail.com

Abstract: Ineffective drug demand management can lead to problems such as imbalanced drug distribution, excess stock, or shortages in community health centers. To address this, data mining can be utilized to support the planning and control process of drug inventory. Clustering techniques were chosen because they are able to group drug data based on certain characteristics, thus identifying stable and unstable drug supply patterns. This study aims to group drug data at Simpang Kawat Community Health Center in Jambi City, which can be used as a reference in planning drug needs in the next period. Data grouping is divided into three categories: slow-moving, medium-moving, and fast-moving. The research data includes attributes of drug name, initial stock, receipt, inventory, usage, and final stock, with a total of 1758 data sets, which were processed using the CRISP-DM framework through the RapidMiner application. Cluster quality evaluation was carried out using the Davies-Bouldin Index (DBI). The results showed that the K-Means algorithm obtained a DBI value of 0.175, smaller than K-Medoids which obtained a value of 0.354. Because a smaller DBI value indicates better cluster quality, K-Means provides more optimal clustering results than K-Medoids. Through these clustering results, community health centers can utilize drug cluster information to support more efficient drug procurement planning, as well as reduce the risk of excess or shortage of stock.

Keywords: data mining; clustering; k-means; k-medoids; davies-bouldin index

Abstrak: Manajemen kebutuhan obat yang tidak efektif dapat menyebabkan permasalahan seperti ketidakseimbangan distribusi obat, kelebihan stok, maupun kekurangan obat di puskesmas. Untuk mengatasi hal tersebut, data mining dapat dimanfaatkan guna mendukung proses perencanaan dan pengendalian persediaan obat. Teknik clustering dipilih karena mampu mengelompokkan data obat berdasarkan karakteristik tertentu sehingga dapat diketahui pola persediaan obat yang stabil maupun tidak stabil. Penelitian ini bertujuan untuk mengelompokkan data obat-obatan di Puskesmas Simpang Kawat Kota Jambi yang dapat digunakan sebagai referensi dalam perencanaan kebutuhan obat pada periode selanjutnya. Pengelompokkan data dibagi menjadi tiga kategori, yaitu lambat habis, sedang, dan cepat habis. Data penelitian mencakup atribut nama obat, stok awal, penerimaan, persediaan, pemakaian, dan stok akhir, dengan jumlah dataset sebanyak 1758 data, yang diproses menggunakan kerangka kerja CRISP-DM melalui aplikasi RapidMiner. Evaluasi kualitas klaster dilakukan menggunakan Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa algoritma K-Means memperoleh nilai DBI sebesar 0,175, lebih kecil dibandingkan K-Medoids yang memperoleh nilai 0,354. Karena nilai DBI yang lebih kecil menunjukkan kualitas klaster yang lebih baik, maka K-Means memberikan hasil klasterisasi yang lebih optimal dibandingkan K-Medoids. Melalui hasil pengelompokan ini, puskesmas dapat memanfaatkan informasi klaster obat untuk mendukung perencanaan pengadaan obat yang lebih efisien, serta mengurangi risiko terjadinya kelebihan maupun kekurangan stok.

Kata kunci: data mining; clustering; k-means; k-medoids; davies-bouldin index

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 733 – 740 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v11i4.4140
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Community Health Centers are first-level health facilities that provide comprehensive, integrated, and sustainable health services to the community and individuals, and play a role as the vanguard in health development through promotive, preventive, curative, and rehabilitative approaches [1], [2]. Ineffective inventory management can result in shortages or excess stock, which can disrupt services and waste resources [3], [4].

To address this issue, data-driven approaches such as data mining can be utilized to identify drug inventory patterns. OSne relevant technique is clustering, which can group data based on specific characteristics, allowing for the identification of fast-, medium-, and slow-to-run drugs [5]. [6].

One of the most commonly used clustering algorithms is K-Means. K-Means is an unsupervised learning algorithm used data into two or more groups [7]. In K-Means, cluster centers are randomly selected according to the number of clusters. Each iteration calculates the membership of the data with respect to the latest centers, and the process stops when the positions of the centers and the membership of the data are stable without change [8]. K-Medoids is similar to K-Means, but divides the data into K clusters with me-doids as cluster centers, which are updated each iteration to accurately represent the cluster members [9].

The selection of K-Means and K-Medoids was based on their high accuracy and ability to efficiently process large amounts of data. Both methods are also flexible, allowing users to determine the number of clusters according to their analysis needs [10].

Based on research conducted by Riva Arsyad Farisa et al. [1] Drug data clustering analysis using K-Means and K-Medoids showed that K-Means had a higher Silhouette Coefficient value (0.627) than K-Medoids (0.536), indicating that K-Means provided better clustering results. Another study conducted by Rahmatika Diana Firdaus et al. [4] A study comparing K-Means and Hierarchical Clustering Single Linkage on drug inventory data showed an evaluation value of 0.8014 for K-Means, while HCC Single Linkage obtained a value of 0.8629, demonstrating the performance of each method.

This study categorizes drug data based on drug type, income, and usage at Simpang Kawat Community Health Center in Jambi City to support stock control decisions. The method used is cluster analysis with the K-Means and K-Medoids algorithms using RapidMiner.

METHOD

This study employed the Cross-Industry Standard Process for Data Mining (CRISP-DM), a six-stage framework from business understanding to knowledge application [11]. The steps to be carried out with CRISP-DM are as shown in Figure 1.

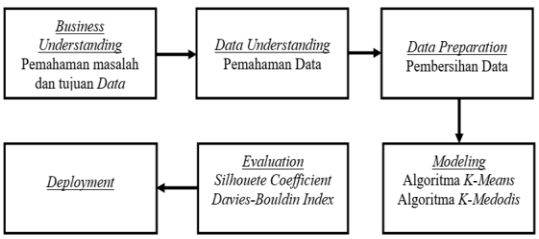

Figure 1. CRISP-DM Research Design

The stages in the Cross-Industry Standard Process for Data Mining (CRISP-DM) method can be explained as follows:

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 733 – 740 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v11i4.4140
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Bussiness Understanding

The Business Understanding stage involves defining the research problem and establishing the objectives of the data mining process, necessitating a comprehensive understanding of both the underlying issues and the intended outcomes.

Data Understanding

The Data Understanding phase emphasizes initial data collection and exploration to identify the fundamental characteristics of the dataset, enabling researchers to comprehend its condition and nature for subsequent analysis.

Data Preparation

The Data Preparation stage includes selecting relevant data and applying preprocessing—such as handling outliers, inconsistencies, missing values, and normalization—to ensure optimal analysis readiness.

Modelling

the Modeling stage, K-Means and K-Medoids clustering are applied to classify drugs into fast-, slow-, and moderately stable categories, with results visualized via cluster diagrams and modeling conducted interactively in RapidMine.

Evaluation

The Evaluation stage assesses clustering quality using the Davies–Bouldin Index (DBI), where lower values denote superior performance, complemented by qualitative analysis of drug groups categorized as fast-moving, slow-moving, or stable :

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}\,(R_{i,j}) \quad (1)$$

In this equation, k denotes the number of clusters, and a smaller non-negative Davies–Bouldin Index (DBI) value indicates better clustering quality..

Deployment

The Deployment stage is the application of the data mining process results into a real environment with the aim that the knowledge or models produced can be used directly to support decision making and improve system performance.

RESULTS AND DISCUSSION

The final results of the research are shown in the form of data clustering through the CRISP-DM stages with the K-Means and K-Medoids algorithms, which are processed using RapidMiner..

Bussiness Understanding

In the Business Understanding phase, the main issue was suboptimal drug stock management, with imbalances between availability and demand—some drugs overstocked and nearing expiry, while others frequently out of stock, affecting care quality.

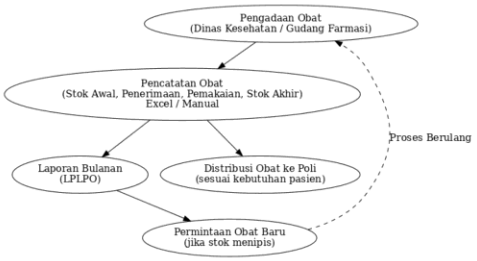

Figure 2 Process Flow for Managing and Recording Drug Stock

Figure 2 illustrates semi-digital medication management at Simpang Kawat Community Health Center, where Excel-based record-keeping that is only administrative and prone to errors can be improved with K-Means and K-Medoids

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 733 – 740 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v11i4.4140
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

clustering to classify medications as fast-, medium-, or slow-acting medications, thereby improving procurement accuracy and operational efficiency.

Data Understanding

In the Data Understanding stage, secondary data from the LPLPO document of Simpang Kawat Community Health Center, Jambi City, covering January 2022 to December 2024 with 1,758 records before cleaning, was used. Exploration revealed missing values, outliers, and imbalances, with some drugs rapidly depleting while others accumulated. The dataset comprised eight attributes drug name, initial stock, receipt, inventory, usage, final stock, optimal stock, and demand as summarized in Table 1.

Table 1 data description

NAME OF DRUG	UNIT	INITIAL STOCK	RECEPTION	SUPPLY	USAGE	FINAL STOCK	OPTIMUM STOCK	REQUEST
2	3	4	5	6 = 4 + 5	7	8 = 7 - 6	9	10

Data Preparation

At this stage, only relevant attributes were retained for modeling, while unit, optimal stock, and per-demand were excluded due to inconsistencies; the final selection is summarized in Table 2.

Table 2 Data Selection Results.

NO	NAME OF DRUG	UNIT	INITIAL STOCK	RECEPTION	SUPPLY	USAGE	FINAL STOCK
1	Acetylcysteine 200 mg	Kaps	10703	13200	23903	4060	19843
2	Acyclovir 200 mg	Tab	0	3000	3000	480	2520
3	Acyclovir 400 mg	Tab	628	0	628	568	60

After attribute selection, missing values were addressed by removing empty columns and records with all zero values, reducing the dataset from 1,758 to 1,279 records as shown in Table 3.

Table 3 Results of handling Missing Value

NO	NAME OF DRUG	INITIAL STOCK	RECEPTION	SUPPLY	USAGE	FINAL STOCK
1	Acetylcysteine 200 mg	10703	13200	23903	4060	19843
2	Acyclovir 200 mg	0	3000	3000	480	2520
3	Acyclovir 400 mg	628	0	628	568	60
....	....	....	....	....	....	....
1277	Hyoscine Butylbromide 10 mg	90	0	90	0	90
1278	Ibuprofen 200 mg	3146	0	3146	1650	1496
1279	Fitomenadion (Vit K) 10 mg	83	500	583	50	533

The next step addresses outliers by removing identified data using RapidMiner operators, thereby cleaning and optimizing the dataset to produce more accurate and relevant clustering results for drug use analysis.

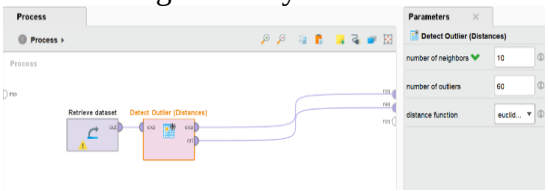

Figure 3 RapidMiner Outlier Detection

Using distance-based k-NN in RapidMiner (10 neighbors, 60 outliers, Euclidean distance), 60 records (5% of 1,279) were identified as outliers and removed, leaving 1,219 records for further analysis (Table 4)

Table 4 Results After Addressing Outliers

NO	INITIAL STOCK	RECEPTION	SUPPLY	USAGE	FINAL STOCK
1	1500	1500	3000	480	2520
2	428	200	628	568	60
3	1650	1500	3150	860	2290
....	....	....	....	....	....
1217	90	0	90	0	90
1218	3146	0	3146	1650	1496
1219	83	500	583	50	533

The next step is data transformation through Min-Max normalization, which linearly scales the original data into a specified value range.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 733 – 740 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v11i4.4140
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

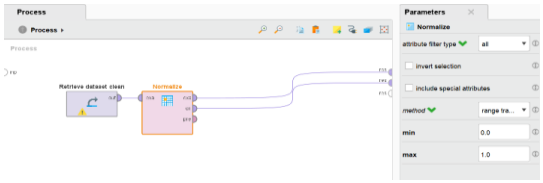

Figure 9 Normalize Min-Max RapidMiner

Normalization was performed using the Min-Max method to scale numeric attributes to a 0–1 range, preventing dominance by any single attribute. The results are presented in Table 5.

Table 5 Min-Max Normalization Results

NO	INITIAL STOCK	RECEPTION	SUPPLY	USAGE	FINAL STOCK
1	0,188	0,094	0,179	0,046	0,230
2	0,054	0,013	0,037	0,055	0,005
3	0,206	0,094	0,188	0,083	0,209
....	....	....	....	....	....
1217	0,011	0,000	0,005	0,000	0,008
1218	0,393	0,000	0,188	0,160	0,136
1219	0,010	0,031	0,035	0,005	0,049

Modelling

K-Means

K-Means clustering was applied with three clusters (fast-, intermediate-, and slow-dissolving drugs) to distinguish stable from unstable items, using Euclidean distance, a maximum of 10 iterations, and the good initial value option Figure 10.

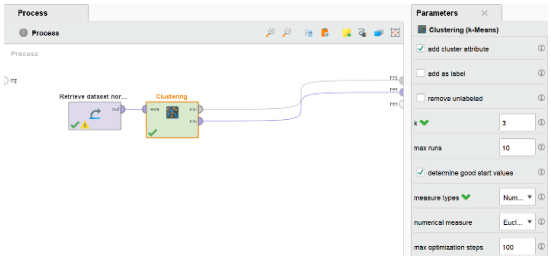

Figure 10 Clustering workflow K-Means RapidMiner

The K-Means process runs until centroids stabilize or the iteration limit is reached, with each centroid representing the characteristics of its drug group based on attribute averages. The centroid results are presented in Table 6.

Table 6 K-Means Centroid Value

Attributes	cluster 0	cluster 1	cluster 2
Initial Stock	0.212	0.016	0.291
Receipt	0.627	0.011	0.117
Supply	0.697	0.019	0.250
Usage	0.499	0.015	0.199
Final Stock	0.593	0.014	0.194

The final K-Means results grouped 26 drugs into cluster 0, 1,059 into cluster 1, and 134 into cluster 2, with most drugs concentrated in cluster 1 and the fewest in cluster 0. The distribution is visualized in a 3D graph Figure 11.

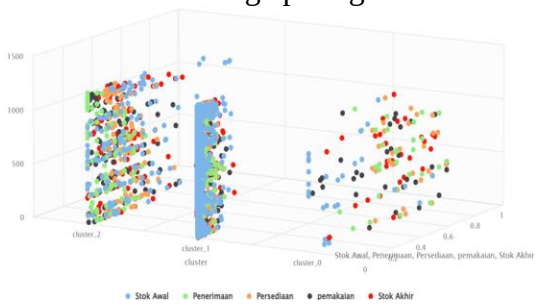

Figure 11 Visualization of K-Means Cluster

The three clusters were then analyzed and labeled as fast-, medium-, or slow-moving drugs based on stock characteristics, with their average attribute percentages compared in Figure 12.

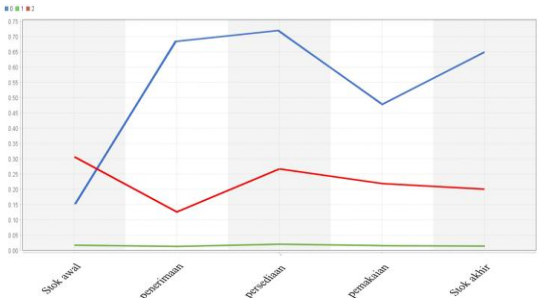

Figure 12 Percentage of Characteristics of Each K-Means Cluster

Figure 12 presents the average attribute patterns of each K-Means cluster, with blue (0), green (1), and red (2) lines

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 733 – 740 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v11i4.4140
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

representing the clusters, and the legend indicating their respective codes:

1. Cluster 0 → High average scores for receipt, inventory, and ending stock, indicating drugs with high availability
2. Cluster 1 → Very low scores across all attributes, representing low-availability or fast-moving drugs.
3. Cluster 2 → Moderate scores, reflecting drugs with relatively stable availability.

K-Medoids

The K-Medoids algorithm was used to classify drugs into fast, medium, and slow-expiratory clusters, selecting medoids as cluster centers for robustness to outliers, with modeling performed in RapidMiner on a normalized dataset (k = 3, 10 iterations, 100 optimization steps) using Euclidean distance, as shown in the workflow diagram.

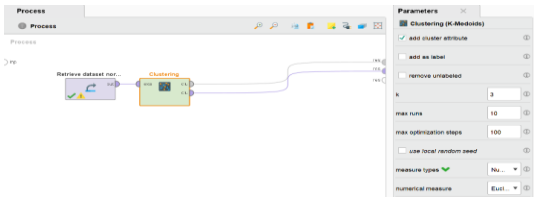

Figure 13 K-Medoids RapidMiner Clustering workflow

In RapidMiner, the K-Medoids operator clustered the data into three groups based on attribute similarities, producing medoid values characterizing each drug group (Table 7), with methods described scientifically to ensure reproducibility and supported by established approaches and relevant references.

Table 7 K-Medoids Centroid Values

Attributes	cluster 0	cluster 1	cluster 2
Initial Stock	0.807	0.010	0.393
Receipt	0.0	0.031	0.0
Supply	0.386	0.035	0.188
Usage	0.180	0.005	0.160
Final Stock	0.419	0.049	0.136

The K-Medoids results grouped 29 drugs into cluster 0, 1,078 into cluster 1, and 134 into cluster 2, with most drugs in cluster 1 and the fewest in cluster 0. The distribution is visualized in a 3D graph (Figure 14).

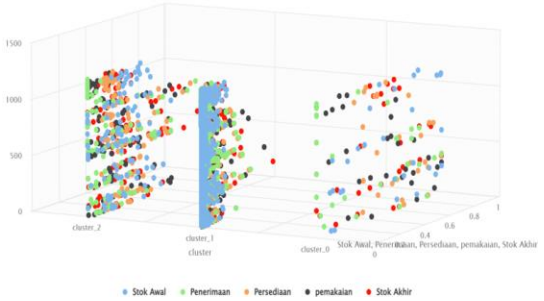

Figure 14 Visualization of K-Medoids Cluster

After forming three clusters, an analysis was conducted to label them as fast-, medium-, and slow-moving drugs based on stock characteristics, with cluster comparisons shown in Figure 15 through average attribute percentages.

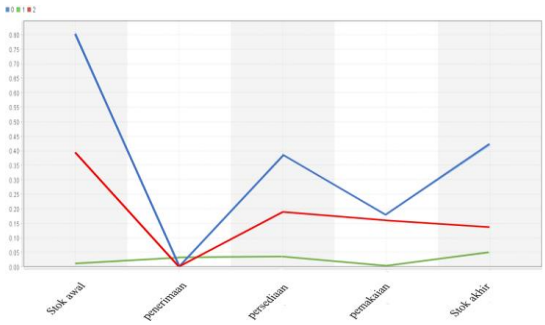

Figure 15 Percentage of Characteristics of Each K-Medoids Cluster

Figure 15 shows the average attribute pattern for each cluster resulting from K-Medoids. The blue (0), green (1), and red (2) lines represent each cluster, while the numbers in the legend in the upper left corner indicate the cluster codes formed:

1. Cluster 0 → High initial stock but low receipt, indicating large starting supplies with minimal replenishment.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 733 – 740 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v11i4.4140
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

2. Cluster 1 → Lowest values across indicators, representing drugs at risk of shortage.
3. Cluster 2 → Moderate and stable inventory and usage, reflecting drugs with balanced availability.

Evaluation

The clustering results were evaluated for validity using the Davies–Bouldin Index (DBI), which compares intra- and inter-cluster distances, where smaller values indicate more compact and well-separated clusters.

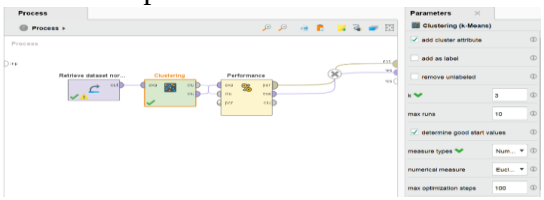

Figure 16 DBI K-Means RapidMiner

Figure 16 illustrates the DBI calculation workflow for K-Means in RapidMiner, where normalized data are clustered with k = 3, up to 10 iterations using Euclidean distance, and evaluated via the Performance (Clustering) operator, yielding a DBI of 0.175, indicative of compact and well-separated clusters.

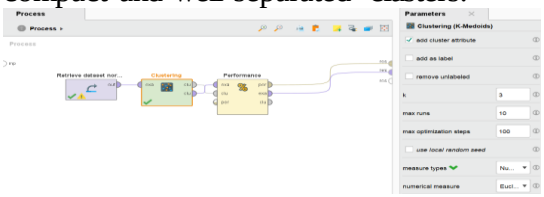

Figure 17 DBI K-Medoids RapidMiner

Figure 17 presents the DBI calculation workflow for K-Medoids in RapidMiner, clustering normalized data with k = 3, 10 iterations, and 100 optimization steps using Euclidean distance, evaluated via the Performance (Clustering) operator, yielding a Davies–Bouldin Index of 0.354, compared with K-Means in Table 8:

Table 8 Evaluation of DBI K-Means and K-Medoids

Algorithm	DBI Value
K-Means	0,175
K-Medoids	0.354

Deployment

K-Means clustering produced three groups: slow-to-discontinue 2.54% (31 drugs), moderate 10.91% (133 drugs), and fast 86.55% (1,055 drugs). These results indicate that most drugs were depleted quickly, as illustrated in Figure 18.

Figure 18 K-Means Deployment Results

K-Medoids clustering grouped 29 drugs (2.38%) as slow-discharge, 112 drugs (9.19%) as moderate-discharge, and 1,078 drugs (88.43%) as fast-discharge, showing that most drugs belong to the fast-discharge cluster Figure 19.

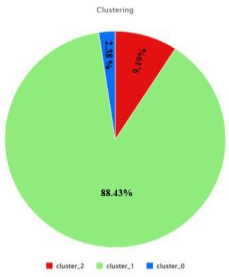

Figure 4.19 K-Medoids Deployment Results

CONCLUSION

This study demonstrates that drug data clustering at Simpang Kawat Community Health Center can be effectively performed using K-Means and K-Medoids. Evaluation results show a significant performance difference, with K-Means achieving a lower Davies–Bouldin Index (0.175) compared to K-Medoids (0.354), indicating superior clustering when data is cleaned, normalized, evenly distributed, and when cluster sizes are relatively balanced. In contrast, K-Medoids is more robust for datasets with many outliers, irregular distributions, or smaller cluster sizes due to its resistance to extreme values. Overall, most drugs fall into the fast-moving category, suggesting inventory management should prioritize replenishment to prevent shortages, and the findings can be integrated into the health center's information system to improve efficiency and minimize waste or drug expiration..

BIBLIOGRAPHY

[1] R. A. Farissa, R. Mayasari, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.

[2] (Permenkes) RI, "Peraturan Menteri Kesehatan (Permenkes) RI Nomor 43 Tahun 2019."

[3] F. Ferlanda, "Implementasi Data Mining Untuk Menentukan Persediaan Stok Obat Di Apotek Enok Menggunakan Metode K-Means Clustering," *JATISI*, vol. 8, no. 3, pp. 1294–1306, 2021, doi: 10.35957/jatisi.v8i3.1066.

[4] S. Michael Alexander Justin Audison Sibaran, I Gede Susrama Mas Diyasa, "Penggunaan K-Means Dan Hierarchical Clustering," vol. 5, no. 2, pp. 1286–1294, 2024.

[5] E. Saha and P. Rathore, "Discovering hidden patterns among medicines prescribed to patients using Association Rule Mining Technique," *Int. J. Healthc. Manag.*, vol. 16, no. 2, pp. 277–286, 2023.

[6] A. Supriyadi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JIPI*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jipi.v6i2.2008.

[7] K. A. Kholil, and R. D. Dana, "Penerapan Data Mining Untuk Clustering Penyakit Diare Menggunakan Algoritma K-Means," *JATI*, vol. 8, no. 3, pp. 3124–3131, 2024.

[8] E. Widodo and W. Hadikristanto, "Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means," *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 408–413, 2023.

[9] S. M. Ramadhan, S. Ramadhani, "Perancangan Website Masyarakat Peduli Sampah Kelurahan Ratu Sima," *J. Penelit. Dan Pengkaj. Ilm. Eksakta*, vol. 1, no. 1, pp. 40–49, 2022, doi: 10.47233/jppie.v1i1.424.

[10] Pelsri Ramadar Noor Saputra and A. Chusyairi, "Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas pada Cakupan Imunisasi Dasar Lengkap," *J. RESTI*, vol. 4, no. 6, pp. 5–12, 2020.

[11] F. Salsabila, I. Fitrianti, and N. Heryana, "Penerapan Metode Crisp-Dm Untuk Analisa Pendapatan Bersih Bulanan Pekerja Informal Di Provinsi Jawa Barat Dengan Algoritma K-Means," *Dinamik*, vol. 28, no. 2, pp. 97–104, 2023, doi: 10.35315/dinamik.v28i2.9454.