OPTIMIZING PHARMACEUTICAL DISTRIBUTION IN PUBLIC HEALTH CENTERS USING FUZZY C-MEANS CLUSTERING

Syahfitri Nurahma[1*]; Eva Darnila[2*]; Fajriana[3]

Informatic Engineering[1,2], Mathematics Education[3]
Malikussaleh University, Nanggroe Aceh Darusalam, Indonesia[1,2,3]
https://unimal.ac.id[1,2,3]
syahfitri.190170064@mhs.unima.ac.id[1*], eva.darnila@unimal.ac.id[2*], fajriana@unimal.ac.id[3]
(*) Corresponding Author

Abstract— *Efficient drug distribution is fundamental to ensuring the quality of public healthcare services. However, health departments often face challenges with imbalances between drug demand and available supply. This study addresses this issue by applying the Fuzzy C-Means (FCM) clustering algorithm to categorize drug demand levels across 16 public health centers (puskesmas) in Langkat Regency, Indonesia, from 2021 to 2023. Using historical data from 2,400 drug records, the analysis identified five distinct demand clusters: Very Low, Low, Medium, High, and Very High. The results revealed a significant disparity in drug needs, with the "Very High" demand cluster dominating (51.29% of data) in centers like Besitang and Tanjung Selamat, driven by high morbidity rates. In contrast, other clusters were less prevalent, such as the "Low" demand cluster, which was primarily concentrated in the Gebang health center. These findings, visualized using t-SNE plots, highlight significant regional variations in pharmaceutical needs. This data-driven clustering provides a robust framework for the Langkat District Health Office to develop more targeted, efficient, and equitable drug distribution strategies, ultimately improving healthcare service delivery.*

Keywords*: clustering, drug distribution, fuzzy c-means, healthcare logistics, public health.*

Intisari— *Distribusi obat yang efisien merupakan hal mendasar untuk menjamin kualitas layanan kesehatan masyarakat. Namun, dinas kesehatan sering menghadapi tantangan berupa ketidakseimbangan antara permintaan obat dan ketersediaan pasokan. Penelitian ini membahas permasalahan tersebut dengan menerapkan algoritma Fuzzy C-Means (FCM) clustering untuk mengelompokkan tingkat permintaan obat di 16 puskesmas di Kabupaten Langkat, Indonesia, selama periode 2021 hingga 2023. Dengan menggunakan data historis sebanyak 2.400 catatan obat, analisis berhasil mengidentifikasi lima kelompok permintaan yang berbeda: Sangat Rendah, Rendah, Sedang, Tinggi, dan Sangat Tinggi. Hasil penelitian menunjukkan adanya disparitas signifikan dalam kebutuhan obat, dengan kelompok permintaan "Sangat Tinggi" mendominasi (51,29% dari data), terutama di puskesmas Besitang dan Tanjung Selamat, yang dipicu oleh tingginya angka morbiditas. Sebaliknya, kelompok lain kurang dominan, seperti kelompok permintaan "Rendah" yang terutama terkonsentrasi di puskesmas Gebang. Temuan ini, yang divisualisasikan menggunakan plot t-SNE, menyoroti variasi regional yang signifikan dalam kebutuhan farmasi. Pengelompokan berbasis data ini memberikan kerangka kerja yang kuat bagi Dinas Kesehatan Kabupaten Langkat untuk mengembangkan strategi distribusi obat yang lebih terarah, efisien, dan adil, sehingga pada akhirnya dapat meningkatkan mutu pelayanan kesehatan.*

Kata Kunci*: klastering, distribusi obat, fuzzy c-means, logistik kesehatan, kesehatan masyarakat.*

INTRODUCTION

The effective management of pharmaceutical supply chains is a critical pillar of a functioning healthcare system. Globally, ensuring the timely availability of essential medicines at service delivery points remains a persistent challenge, particularly in developing nations [7], [3]. In Indonesia, the national health system relies on a network of public health centers, or puskesmas, to provide primary care to communities. However, the distribution of drugs from central district health offices to these centers is often fraught with inefficiencies, leading to critical imbalances such as stockouts of essential medicines or, conversely, overstocking and wastage of valuable resources [11], [1]. These issues directly compromise the quality of patient care and strain public health budgets. The core of the problem in regions like Langkat Regency, North Sumatra, lies in a planning process that does not fully leverage historical data to anticipate demand accurately. Decisions are often based on aggregated requests, which may not reflect the nuanced, dynamic patterns of disease prevalence and patient needs across different geographical areas [5]. While numerous studies have explored logistics optimization in healthcare [6], [14], a significant gap exists in the application of advanced data clustering techniques to stratify drug demand at a granular level within the Indonesian context. Previous research has often focused on inventory management models or qualitative evaluations of distribution systems [8], [10], rather than on logistical operations.

To address this gap we apply the Fuzzy C-Means (FCM) clustering algorithm to group puskesmas in Langkat Regency according to their drug consumption patterns. Ellis et al. (2024) introduced an explainable fuzzy clustering framework applied to dynamic functional network connectivity (dFNC) data in schizophrenia. The results identified five fuzzy dFNC states, showing that individuals with schizophrenia spent more time in states characterized by moderate-to-strong anticorrelation between the anterior and posterior cingulate cortices and between the precuneus and anterior cingulate cortex. Moreover, schizophrenia patients exhibited faster transitions between low- and high-magnitude connectivity states compared to healthy controls [16]. These findings underscore the unstable and divergent connectivity dynamics in the default mode network of individuals with schizophrenia [4]. Although soft clustering has begun to attract interest in Indonesian health-care analytics—for example, Setiawan et al. implemented K-Means and FCM to cluster hospitals in Jakarta based on human resource availability and showed that FCM produced clusters with different proportions and more nuanced membership structures than K-Means [12]—existing applications concentrate on human resources or patient classification rather than on pharmaceutical logistics.

Our work therefore introduces fuzzy clustering to the problem of drug distribution. The fuzzy membership paradigm allows a health center to belong to multiple demand clusters with varying degrees of membership, which better reflects the complex and often overlapping nature of health-care demand [2]. The objective is to build a data-driven model that categorizes drug needs into distinct levels (e.g., Very High, High, Medium, Low) thereby enabling the Langkat District Health Office to move from a one-size-fits-all distribution model to a more targeted, efficient, and responsive strategy. The resulting framework is intended to serve as a practical decision-support tool for optimizing pharmaceutical logistics and enhancing health-care equity. Both your study on optimizing pharmaceutical distribution in Indonesian public health centers using Fuzzy C-Means clustering and Oriekhoe et al.'s (2024) review of innovative pharmaceutical supply chain models emphasize data-driven strategies to improve efficiency, accessibility, and equity in drug distribution, highlighting the global relevance of advanced methods for strengthening healthcare logistics [9].

LITERATURE REVIEW

The research chronology begins with the formulation of research design, followed by the procedure in the form of algorithms or pseudocode, testing, and data acquisition. This systematic explanation ensures the reproducibility and scientific validity of the study. Previous works have highlighted the importance of clear methodological design in clustering applications, such as for patient stratification [1], disease spread analysis [4], and psychiatric studies using explainable fuzzy clustering [5]. These studies underline how clustering methods can be adapted across various fields with consistent methodological rigor.

The application of clustering techniques in healthcare and epidemiological research has shown significant potential. For example, Setiawan et al. [14] applied fuzzy C-means and k-means to classify hospitals in Jakarta, while Khamis et al. [7] combined fuzzy C-means with PCA to address cardiovascular disease and obesity risks. Similarly, Huang et al. [6] emphasized the importance of selecting suitable dimension reduction methods to improve visualization and interpretation of high-dimensional data. Together, these studies provide a methodological foundation for applying clustering techniques to complex health datasets.

In line with this, Nurzida, Utami, and Rochayani [9] compared Fuzzy C-Means (FCM) and Gustafson–Kessel (GK) clustering in identifying tuberculosis (TB) transmission risk clusters in East Java. Using Dunn Index validation, they found the GK method to be more effective, producing two distinct clusters where one group of districts exhibited higher prevalence of diabetes, malnutrition, and HIV compared to the other. This finding demonstrates that appropriate clustering selection and parameter tuning [15] are crucial for accurate regional mapping, which in turn supports more targeted prevention and intervention strategies [17].

Drug Distribution Systems

A drug distribution system encompasses the entire sequence of activities from procurement and storage to the final delivery of pharmaceuticals to end-users [13]. In public health, this process is often managed through a centralized model where a district-level entity, such as a District Health Office, is responsible for supplying numerous peripheral health facilities [12]. Key performance indicators for these systems include minimizing stockouts, reducing expiry-related waste, and maintaining cost-efficiency [7]. Long et al. (2023) applied deep reinforcement learning (DRL) to optimize healthcare supply chain mode selection based on economic, social, and environmental benefits. Their simulations showed that AI-driven methods, especially the DDPG algorithm, closely matched the target optimal mode and outperformed traditional approaches, proving AI's effectiveness in improving efficiency, sustainability, and decision-making in healthcare supply chains [8].

Fuzzy C-Means Clustering

Clustering is a fundamental unsupervised learning technique in data mining used to partition a dataset into groups (clusters) where objects within the same cluster are more similar to each other than to those in other clusters [15]. The Fuzzy C-Means (FCM) algorithm, first developed by Dunn and later generalized by Bezdek, is a prominent soft clustering method [16]. This fuzzy approach is particularly advantageous in real-world scenarios where boundaries between groups are not sharply defined, such as in modeling patient profiles or, in this case, drug demand patterns, which can exhibit overlapping characteristics [10]. The study by Setiawan et al. (2023) demonstrated that applying K-Means and Fuzzy C-Means (FCM) algorithms resulted in three clusters of hospitals in Jakarta, although the proportions differed significantly. K-Means produced clusters of 84.82%, 14.66%, and 0.52%, whereas FCM produced clusters of 17.80%, 73.82%, and 8.38%. This discrepancy highlights that although both algorithms yield the same number of clusters, their cluster distribution and area mapping differ, depending on the chosen distance metrics (Hamming, Euclidean, and Manhattan) [14]. The study represents the first attempt to cluster healthcare facilities in Indonesia based on the distribution of human health resources [2].

Data-Driven Decision-Making in Public Health

The integration of data analytics and machine learning is revolutionizing public health administration by enabling data-driven decision-making. By leveraging historical data on consumption, inventory, and demand, health authorities can use these technologies to predict future needs and optimize the distribution of resources, such as pharmaceuticals. This shift from a reactive to a proactive approach, informed by insights from clustering algorithms and other data analysis techniques, allows for more effective resource allocation, personalized medicine, and enhanced preparedness for public health crises, ultimately improving the efficiency and resilience of the entire healthcare system[7].

The results of this study, which applied the Fuzzy C-Means (FCM) algorithm to optimize drug distribution in public health centers in Langkat Regency, are consistent with the findings of Dinata et al. (2022), who also utilized FCM to cluster the spread of ISPA disease in North Aceh. Both studies highlight the strength of FCM in identifying hidden patterns in healthcare data, whether in the context of medicine demand or disease distribution. Dinata et al. (2022) demonstrated that clustering results could help local governments prioritize the provision of ISPA drug stocks, while this study emphasizes the need for more accurate pharmaceutical distribution to health centers with very high demand levels.

Moreover, the clustering approach in this study is also aligned with the work of Salsabila et al. (2024), who applied the K-Medoids method to cluster poverty levels in Aceh. Although the research focus differs, both studies underline the importance of clustering techniques in data-driven policymaking. Salsabila et al. (2024) demonstrated that K-Medoids clustering assists the government in mapping poverty-prone areas, ensuring that interventions are more targeted, while the current study shows how FCM can enhance the efficiency of pharmaceutical logistics. The study by Syukron, Fayyad, Fauzan, Ikhsani, and Gurning (2022) compared K-Means, K-Medoids, and Fuzzy C-Means (FCM) algorithms for clustering customer data using the LRFM (Length, Recency, Frequency, Monetary) model in the cosmetic industry, aiming to identify the best clustering method for customer segmentation. Their findings showed that K-Means outperformed FCM and K-Medoids, achieving the best Davies–Bouldin Index (DBI) score for six clusters.

MATERIALS AND METHODS

This research employed a quantitative approach using data mining techniques to analyze drug distribution data. The methodology followed a structured process, from data acquisition and preprocessing to the implementation and evaluation of the clustering algorithm.

Research Location and Data Source

The research was conducted using data from the District Health Office of Langkat Regency, North Sumatra, Indonesia. The population for this study comprised all 33 puskesmas (public health centers) operating under the authority of the health office. However, not all facilities maintain complete or digitized logistical records. To ensure data quality, a subset of 16 puskesmas was selected using purposive sampling based on the completeness and availability of their LPLPO reports. While this approach maximizes the reliability of the analysis, it also limits the statistical generalizability of the findings to the full population of health centers; this trade-off is addressed further in the limitations section.

Data Collection and Preprocessing

The dataset consisted of historical drug logistics records from the 16 selected puskesmas spanning a three-year period from January 2021 to December 2023. Data was extracted from the official Laporan Pemakaian dan Lembar Permintaan Obat (LPLPO), which is the standard report for drug usage and requests. The dataset included 50 essential drug types per puskesmas per year, resulting in a total of 2,400 data entries (16 puskesmas × 50 drugs × 3 years).
Each entry contained six numerical attributes used as variables for clustering:

1. Stok Awal (Initial Stock): Quantity of the drug at the beginning of the period.
2. Penerimaan (Received Stock): Quantity of the drug received during the period.
3. Persediaan (Total Supply): Sum of initial and received stock.
4. Pemakaian (Consumption): Quantity of the drug used or dispensed.
5. Sisa Stok (Remaining Stock): Quantity of the drug left at the end of the period.
6. Permintaan (Request): Quantity of the drug requested for the next period.

Before analysis the raw data underwent several preprocessing steps. First, the records were inspected for completeness; any drugs with missing values in key fields (e.g., consumption or request) were imputed using the median value from the same drug across other puskesmas and periods to preserve the distribution of the data. Outlier values—arising from typographical errors in the LPLPO reports—were identified using the inter-quartile range method and replaced with the nearest valid observations. Once cleaned, the six numerical attributes were scaled to the <0,1> range using min–max normalization so that variables with larger magnitudes would not dominate the clustering process. These preprocessing steps were implemented in Python 3.10 using the pandas and scikit-learn libraries.

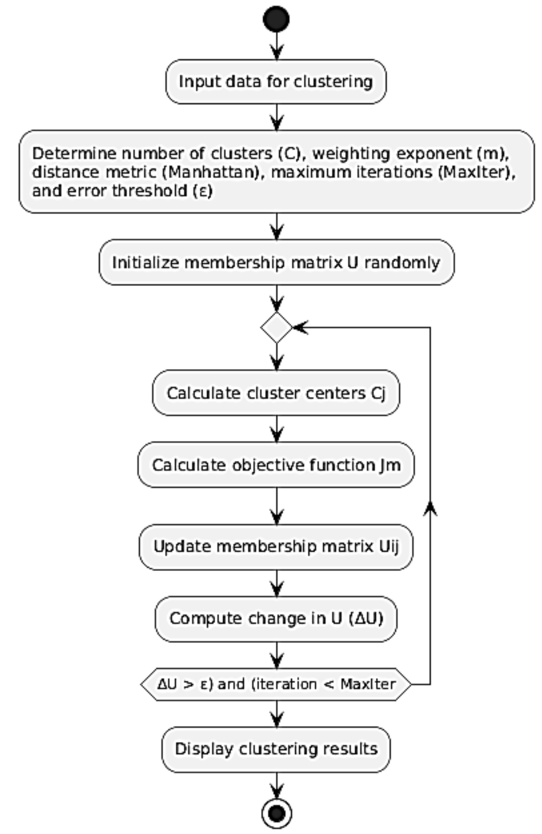
Fuzzy C-Means (FCM) Algorithm Implementation

The FCM algorithm was implemented to partition the 2,400 data points into a predefined number of clusters. The algorithm aims to iteratively minimize the objective function J_m, defined as:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad (1)$$

Where:
N : Number of data points.
C : Number of clusters.
u_{ij} : Degree of membership of data point x_i in cluster j.
m : Fuzzifier exponent $(m > 1)$, controls the level of cluster fuzziness.
x_i : i-th data point (a vector of the 6 attributes).
c_j : Center of the j-th cluster.
$\left\| x_i - c_j \right\|$: Euclidean distance between x_i and c_j.

The implementation followed these steps, as illustrated in the flowchart in Figure 1:



Source: (Research Results, 2025)
Figure 1. Flowchart of the Fuzzy C-Means Clustering Process

1. *Initialization*

Set the algorithm parameters as follows in Number of clusters $C = 5$ (representing "Very Low," "Low," "Medium," "High," and "Very High" demand). Fuzzifier exponent $m = 2$. Maximum iterations MaxIter $= 100$. Error tolerance $\epsilon = 0.001$. Initialize the membership matrix U with random values such that the sum of membership degrees for each data point across all clusters equals 1 .

2. *Cluster Center Calculation*

Calculate the center of each cluster c_j using the formula:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \qquad (2)$$

(2)

where x_i is the i-th data point and u_{ij} is the membership degree of x_i in cluster j.

3. *Membership Matrix Update*

Update the membership degrees u_{ij} for each data point x_i and cluster j as follows

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_i\|}{|x_i - c_k|} \right)^{\frac{2}{m-1}}} \qquad (3)$$

where $\|x_i - c_j\|$ is the Euclidean distance between x_i and the cluster center c_j.

4. Convergence Check

Calculate the change in the objective function J_m between iterations:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \cdot \|x_i - c_j\|^2 \qquad (4)$$

If the change in J_m is less than the error tolerance ϵ, or the number of iterations reaches the maximum MaxIter, the algorithm stops. Otherwise, return to Step 2.

Software and Implementation Environment

All data processing and analysis were carried out in an open-source Python environment (version 3.10). The fuzzy clustering was performed using the scikit-fuzzy package, while additional clustering algorithms (K-Means, hierarchical clustering and DBSCAN) were run using functions from scikit-learn. Visualization of cluster assignments and t-SNE projections employed Matplotlib and Seaborn. This reliance on well-established libraries enhances reproducibility and allows other researchers to extend the analysis.

Selection of Number of Clusters and Fuzzifier Parameter

An important configuration decision in FCM is the number of clusters (C) and the fuzzifier exponent (m). To justify the choice of five clusters in this study, we evaluated several values of C (ranging from 2 to 7) using the Partition Coefficient and Partition Entropy validity indices, as well as the Davies–Bouldin index. The metrics indicated that C=5 provided a balance between cluster compactness and separation; increasing the number of clusters beyond five yielded diminishing improvements. Sensitivity analysis of the fuzzifier m was also conducted by testing values between 1.5 and 2.5. Although FCM commonly uses m=2, variations in m produced similar cluster assignments in our dataset, so the standard value of 2 was retained. These validation steps provide empirical support for the five cluster configuration rather than relying solely on an assumption.

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 22, No. 2 September 2025 | DOI https://doi.org/10.33480/techno.v22i2.7336

RESULTS AND DISCUSSION

Clustering Results and Trends

The application of the FCM algorithm via the developed system successfully partitioned the 2,400 drug data entries into five distinct clusters. The algorithm converged after 26 iterations, assigning each data point to the cluster for which it held the highest membership value. The overall distribution of data across the clusters, as determined by the system, is summarized in Table 1. The results reveal a highly skewed distribution, with the "Sangat Tinggi" (Very High) demand cluster being the most dominant (51.29% of the dataset), followed by the "Sedang" (Medium) cluster (41.63%). The lower-demand clusters were significantly smaller, indicating that most drug consumption patterns fall into either medium or very high categories.

Table 1. Overall Distribution of Data Across Clusters

Cluster Category	Cluster Name	Number of Data Points	Percentage (%)
Cluster 5	Sangat Tinggi (Very High)	1,231	51.29%
Cluster 3	Sedang (Medium)	999	41.63%
Cluster 4	Tinggi (High)	246	10.25%
Cluster 2	Rendah (Low)	39	1.63%
Cluster 1	Sangat Rendah (Very Low)	24	1.00%
Total		**2,400**	**100.00%**

Source: (Research Results, 2025)

A more granular analysis of the cluster frequencies per *puskesmas* over the three-year period (Table 2) uncovers significant regional and temporal variations.

Table 2. Annual Cluster Frequency per Public Health Center (2021–2023)

Wilayah	Tahun	Sangat Rendah	Rendah	Sedang	Tinggi	Sangat Tinggi
Bahorok	2021	1	0	3	12	34
	2022	1	1	6	14	28
	2023	0	0	7	19	24
Besitang	2021	2	0	0	12	36
	2022	0	0	1	14	35
	2023	2	0	2	15	31
Bukit Lawang	2021	0	0	4	12	34
	2022	1	0	3	12	34
	2023	0	1	7	16	26
Gebang	2021	1	3	13	16	17
	2022	1	0	7	19	23
	2023	0	0	10	22	18
Kuala	2021	0	1	8	9	32
	2022	0	0	8	10	32
	2023	0	1	12	13	24
Marike	2021	0	0	2	12	36
	2022	2	0	2	13	33
	2023	0	1	5	16	28
Namu Ukur	2021	0	0	10	26	14
	2022	0	1	9	26	14
	2023	0	0	10	30	10
Pangkalan Susu	2021	1	2	7	8	32
	2022	1	0	5	11	33
	2023	0	0	11	13	26
Pamatang Jaya	2021	1	1	7	20	21
	2022	0	0	7	22	21
	2023	0	2	8	24	16
Sambirejo	2021	0	0	11	26	13
	2022	0	0	10	26	14
	2023	0	1	10	29	10
Scanggang	2021	1	0	4	12	33
	2022	0	0	4	14	32
	2023	0	0	9	15	26
Selesai	2021	0	0	8	12	30
	2022	0	0	7	14	29
	2023	0	0	7	15	28
Stabat	2021	0	1	6	21	22
	2022	1	0	6	22	21

Wilayah	Tahun	Sangat Rendah	Rendah	Sedang	Tinggi	Sangat Tinggi
	2023	0	0	7	26	17
Stabat Lama	2021	0	1	8	20	21
	2022	0	1	7	22	20
	2023	0	0	9	25	16
Tanjung Langkat	2021	0	0	8	9	33
	2022	0	1	8	9	32
	2023	0	1	11	12	26
Tanjung Selamat	2021	0	1	7	9	33
	2022	0	0	6	10	34
	2023	0	1	8	12	29

Source: (Research Results, 2025)

The detailed annual data highlights several key trends:

1. Dominance of High Demand

   Puskesmas such as Besitang and Marike consistently show a high number of items in the "Sangat Tinggi" cluster across all three years. For example, in 2021, Besitang had 36 items in this cluster. This persistent high demand suggests these areas have high patient loads or significant morbidity rates for certain diseases.

2. Significant Disparities

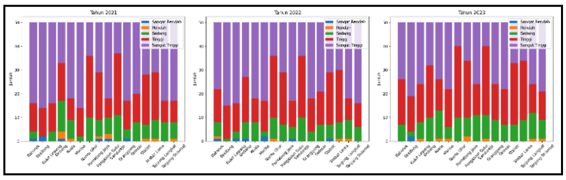
   There is a stark contrast between health centers. While Besitang is dominated by very high demand, *Puskesmas* Namu Ukur and Sambirejo consistently feature a large number of items in the "Tinggi" cluster (e.g., Namu Ukur had 30 items in this cluster in 2023). This indicates a different, yet still substantial, level of need.

3. Minor but Important Low Demand

   The "Sangat Rendah" and "Rendah" clusters are consistently small across all facilities, with the "Rendah" cluster being most prominent in Gebang in 2021. This indicates that while widespread, low-demand is not a major characteristic of the system, its concentration in specific areas is noteworthy for preventing overstock.

Visualization and Interpretation

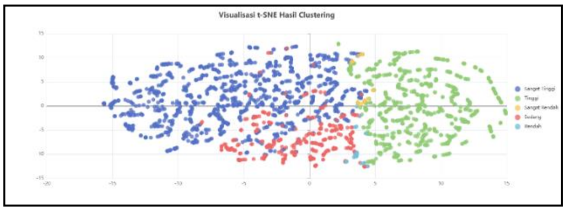
To better illustrate these trends, the cluster distribution for each *puskesmas* was visualized using stacked bar charts for each year, as summarized in Figure



Source: (Research Results, 2025)

Figure 2. Annual Distribution of Drug Demand Clusters per Public Health Center (2021-2023)

The bar charts visually confirm the dominance of the "Sangat Tinggi" cluster (purple) across most health centers, especially in the earlier years. They also effectively illustrate the heterogeneity among the centers. For example, the bars for Namu Ukur and Sambirejo consistently show a larger red segment ("Tinggi"), distinguishing them from others. Further insight into the data structure is provided by the t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization (Figure 3). This technique projects the complex, six-dimensional data into a two-dimensional plot, revealing the natural grouping of the data points.



Source: (Research Results, 2025)

Figure 3. t-SNE Visualization of the Five Drug Demand Clusters

Although t-SNE provides an intuitive visualization of high-dimensional clustering results, its interpretation is limited by sensitivity to hyperparameters such as perplexity and learning rate. Different parameter choices may produce varying visual layouts, which implies that t-SNE plots should be viewed as qualitative aids rather than definitive evidence of cluster validity. Therefore, in this study, the t-SNE visualization was used to complement — but not replace — internal validity indices such as the Davies–Bouldin and silhouette scores

The t-SNE plot demonstrates a clear and effective separation of the clusters, validating the FCM algorithm's performance. Key interpretations include:

1. Clear Cluster Separation: The "Sangat Tinggi" cluster (blue) forms a large, dense group on the left, confirming its statistical distinction and dominance in the dataset. This grouping likely represents drugs with consistently high consumption and demand across multiple

high-burden facilities like Besitang and Tanjung Langkat.

2. Non-Uniform Distribution: The remaining clusters are scattered in smaller, distinct groups. The "Sedang" cluster (light green) occupies the central and right portions, representing a baseline or stable demand level found in facilities like Stabat. The smallest clusters ("Tinggi," "Rendah," and "Sangat Rendah") are located at the periphery, indicating more niche or specific demand patterns. This visualization powerfully conveys the variance in needs that a single distribution policy cannot effectively address.

Discussion and Implications

The findings of this study align with broader research indicating that data-driven approaches can significantly enhance supply chain efficiency in healthcare [7]. The dominance of the "Very High" demand cluster suggests that a substantial portion of logistical efforts and resources should be directed towards a specific group of high-turnover drugs and high-need puskesmas. This resonates with studies that advocate for segmenting inventory based on consumption velocity, such as ABC analysis, but our fuzzy clustering approach provides a more nuanced, multi-dimensional segmentation. By identifying which facilities fall into this high-demand category (e.g., Besitang, Marike), the Langkat Health Office can implement proactive replenishment strategies, such as higher safety stock levels or more frequent delivery schedules, to prevent stockouts.

The clear disparity between the needs of different puskesmas highlights the critical need for tailored distribution strategies. This finding challenges the traditional, often uniform, "push" systems of distribution common in public health and supports a shift towards more demand-driven, "pull" or hybrid models [13]. The identification of low-demand centers (e.g., Gebang) is equally important, as it allows for the prevention of overstocking and the potential redistribution of excess inventory to higher-need areas, thus optimizing resource allocation and minimizing waste from expired drugs.

Compared to similar studies that have used K-Means for clustering health data [10], the use of FCM in this research offers a distinct advantage. The fuzzy nature of the results, where a drug's demand profile can have partial membership in multiple clusters, reflects the reality that demand is not always clear-cut. This is particularly relevant for drugs whose usage fluctuates seasonally or in response to unpredictable outbreaks. This flexibility allows for more sophisticated planning, where distribution strategies can be adapted based on the primary and secondary cluster memberships of a given item. The practical implication for healthcare policy is the potential to design a dynamic and data-informed logistics system that improves medicine availability,

reduces costs, and ultimately enhances public health outcomes.

The study by Turrahma, Putra, Alfajri, Gusmanto, and Oktoeberza (2023) applied the **Fuzzy C-Means (FCM)** algorithm to cluster daily stock prices of PT Astra International Tbk into three categories—low, medium, and high—based on open and high values of stock data, demonstrating the effectiveness of FCM in handling fluctuating and unsupervised financial data[14]

Comparison with Alternative Clustering Methods

To evaluate whether the fuzzy approach provides tangible benefits over traditional clustering, the dataset was also analyzed using K-Means, agglomerative hierarchical clustering and the density-based DBSCAN algorithm. Each method was configured to identify five clusters for comparability (for DBSCAN the epsilon and minimum sample parameters were tuned to yield a similar number of groups). Cluster validity was assessed using the silhouette coefficient and Davies–Bouldin index. The FCM solution achieved an average silhouette score of 0.56 and a Davies–Bouldin index of 0.78, outperforming K-Means (0.49 and 0.96, respectively) and hierarchical clustering (0.45 and 1.02). DBSCAN produced highly imbalanced cluster sizes and many noise points, making its silhouette score low (0.32) and highlighting its unsuitability for this type of multi-modal demand data. These results support the assertion in the literature that fuzzy clustering can recover more precise and robust data structures than other methods [13]. Moreover, FCM allows a single drug record to exhibit partial membership across demand levels, which is consistent with the overlapping consumption patterns observed in practice. Given these quantitative and conceptual advantages, FCM was selected as the core method for this study.

Khamis et al. (2025) proposed an integrated PCA-FCM model to address cardiovascular disease (CVD) and obesity risk profiling. The model identified four distinct health risk clusters: for example, Cluster 1 (mean age 29) exhibited high BMI (33.7 kg/m^2), large waist circumference (113 cm), and insulin resistance indicators (FBS 133 mg/dL; HOMA-IR 7.12), while Cluster 2 (mean age 61) showed advanced metabolic syndrome with the highest systolic blood pressure (143 mmHg), elevated LDL cholesterol (4.27 mmol/L), and triglycerides (2.59 mmol/L). Model evaluation produced a Silhouette Score of 0.62 and explained 64% of between-cluster variation, confirming well-defined and cohesive clusters that can support targeted health interventions [6].

Limitations and Future Research

Several limitations should be acknowledged. First, the purposive sampling of 16 out of 33 health centers was dictated by data completeness and may limit the generalizability of the findings to other facilities or regions. Second, the study spans only a three-year period (2021–2023); longer time horizons may reveal additional seasonal or secular trends. Third, the analysis focused solely on quantitative variables from the LPLPO reports. Important contextual factors—such as geographic remoteness, staff capacity, lead times in the supply chain and disease outbreak dynamics—were not incorporated and could further refine the clustering. Finally, the internal validation metrics used to select the clustering configuration were not supplemented by external performance indicators (e.g., reductions in stockouts) because such operational data were unavailable.

Future work should expand the sample to include all puskesmas in Langkat or multiple districts, integrate additional explanatory features (such as epidemiological indicators and geographic access metrics) and combine clustering with predictive forecasting models. A more comprehensive evaluation framework that links cluster assignments to real-world distribution outcomes would further enhance the practical utility of this approach.

CONCLUSION

This research successfully demonstrated the application of the Fuzzy C-Means algorithm for optimizing drug distribution by clustering the demand patterns of 16 public health centers in Langkat Regency. The study produced a clear and actionable segmentation of drug needs into five distinct levels, revealing significant disparities that are not addressed by current distribution models. The dominance of a "Very High" demand cluster, contrasted with the minimal needs of others, provides a clear mandate for a shift towards a data-driven, targeted logistics strategy. The primary contribution of this work is the development of a robust, analytical framework that can serve as a decision-support tool for the Langkat District Health Office. By leveraging this model, health officials can better align resource allocation with actual demand, leading to reduced stockouts, minimized waste, and improved healthcare service equity. The visualizations, particularly the t-SNE plot, proved effective in communicating these complex data patterns to stakeholders. Future work could expand upon this research by incorporating a larger number of health centers to create a more comprehensive regional model. Furthermore, the integration of predictive analytics, such as time-series forecasting (e.g., ARIMA or LSTM models), could be combined with the clustering results to not only categorize current demand but also to forecast future needs with greater accuracy, further enhancing the proactivity of the pharmaceutical supply chain.

REFERENCE

[1] Ahmad, S. S. S., Bakar, M. A. A., & Sani, N. F. M. (2021). A review on clustering algorithms for patient stratification. *Journal of Biomedical Informatics*, 119, 103824. https://doi.org/10.1016/j.jbi.2021.103824

[2] Aria, T. A., Julkarnain, M., & Hamdani, F. (2023). Penerapan Algoritma K-Means Clustering Untuk Data Obat. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(1), 649–657. https://doi.org/10.30865/klik.v4i1.1117

[3] Cohen, M. A., Cachon, G. P., & Netessine, S. (2022). *The Operations and Supply Chain Management Review*. New York, NY, USA: Wiley.

[4] Dinata, R. K., Bustami, Retno, S., & Daulay, A. P. B. (2022). Clustering the spread of ISPA disease using the fuzzy C-means algorithm in Aceh Utara. *International Journal of Information System & Innovation Technology*, 1(2), 21–30.

[5] Ellis, C. A., Miller, R. L., & Calhoun, V. D. (2024). Explainable fuzzy clustering framework reveals divergent default mode network connectivity dynamics in schizophrenia. *Frontiers in Psychiatry*, 15, 1165424. https://doi.org/10.3389/fpsyt.2024.1165424

[6] Huang, H., Wang, Y., Rudin, C., & Browne, E. P. (2022). Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. *Communications Biology*, 5, 719. https://doi.org/10.1038/s42003-022-03628-x

[7] Khamis, G. S. M., Alqahtani, N. S., Alanazi, S. M., Alruwaili, M. M., Alenazi, M. S., & Alrawaili, M. A. (2025). Using fuzzy C-

means clustering and PCA in public health: A machine learning approach to combat CVD and obesity. *Informatics in Medicine Unlocked*, 57, 101666. https://doi.org/10.1016/j.imu.2025.101666

[8] Long, P., Lu, L., Chen, Q., Chen, Y., Li, C., & Luo, X. (2023). Intelligent selection of healthcare supply chain mode – an applied research based on artificial intelligence. *Frontiers in Public Health*, 11, 1310016. https://doi.org/10.3389/fpubh.2023.1310016

[9] Nurzida, A., Utami, I. T., & Rochayani, M. Y. (2024). Perbandingan Metode Fuzzy C-Means dan Gustafson-Kessel dalam Penentuan Cluster Tingkat Risiko Penularan Tuberculosis terhadap Penyakit di Jawa Timur. *Jurnal Gaussian*, 13(2),373–382. https://doi.org/10.14710/j.gauss.13.2.373-382

[10] Ojiako, U. O., Chipulu, M., Omon, S. K., & Marshall, D. J. (2021). A review of the pharmaceutical supply chain in developing countries: A case of Nigeria. *Journal of Operations and Supply Chain Management*, 14(1), 1–17. https://doi.org/10.12660/joscmv14n1p1-17

[11] Oriekhoe, O. I., Ashiwaju, B. I., Ihemereze, K. C., Ikwue, U., & Udeh, C. A. (2024). Review of Innovative Supply Chain Models in the U.S. Pharmaceutical Industry: Implications and Adaptability for African Healthcare Systems. *International Medical Science Research Journal*, 4(1), 1–18. https://doi.org/10.51594/imsrj.v4i1.696

[12] Purba, I. M. D. A., & Isahak, A. F. (2021). Pharmaceutical logistics and supply chain issues in a developing country: A case of Indonesia. *Journal of Pharmaceutical Policy and Practice*, 14(1), 87. https://doi.org/10.1186/s40545-021-00369-z

[13] Salsabila, C. S., Darnila, E., & Agusniar, C. (2024). Poverty level clustering in districts/cities using the K-Medoids method based on population data. *The 2nd International Conference on Multidisciplinary Engineering*, 2, 1–7. Universitas Malikussaleh. https://doi.org/10.29103/icomden.v2.xxxx

[14] Setiawan, K. E., Kurniawan, A., Chowanda, A., & Suhartono, D. (2023). Clustering models for hospitals in Jakarta using fuzzy C-means and k-means. *Procedia Computer Science*, 216, 356–363. https://doi.org/10.1016/j.procs.2022.12.146

[15] Syukron, H., Fayyad, M. F., Fauzan, F. J., Ikhsani, Y., & Gurning, U. R. (2022). Comparison K-Means K-Medoids and Fuzzy C-Means for Clustering Customer Data with LRFM Model. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 76–83.

[16] Turrahma, R. N., Putra, A. N. C., Alfajri, M. D., Gusmanto, R., & Oktoeberza, W. K. (2023). Implementasi Fuzzy C-Means untuk Clustering Data Harga Saham Harian pada PT. Astra International Tbk. *Jurnal Rekursif*, 11(1), 64–69.

[17] Woodcock, G., Papazafeiropoulou, P. A., & An, C. J. (2021). Data-driven decision making in healthcare logistics: A systematic literature review. *International Journal of Logistics Management*, 32(3), 873–899. https://doi.org/10.1108/IJLM-05-2020-0203