Journal of Science and Social Research ISSN 2615 – 4307 (Print)
April 2026, IX (2): 2581 – 2587 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

OPTIMIZING PATIENT SERVICE PRIORITY DETERMINATION IN THE EMERGENCY DEPARTMENT USING THE K-MEANS ALGORITHM AT RSUD LANGSA ACEH

Salwa Nur JB[1], Muhammad Irfan Sarif[2], Lola Astri Nadita[3], Fachrurazy[4], Lewika Tampubolon[5]
Universitas Pembangunan Panca Budi, Medan
e-mail: [1]salwanur2015@gmail.com, [2]irfanberbagi@gmail.com, [3]lolanadita@gmail.com, [4]oxxy.keren@gmail.com, [5]lewikatampubolon87@gmail.com

Abstract: *Emergency Department (ED) services require speed, accuracy, and coordination among medical staff. Langsa Aceh Regional General Hospital, a Type B hospital, faces a high volume of patient visits that could lead to inefficiencies, particularly in determining service priorities. This study aims to optimize ED patient segmentation to support service prioritization using the K-Means clustering algorithm. The data used consists of 31,761 ED patient records from Langsa Regional General Hospital in 2025; after preprocessing, 24,062 records meeting the analysis criteria were obtained. Research variables included visit frequency, visit interval, type of service, duration of service, and patient urgency level. The research method employed a quantitative approach using data mining techniques with the K-Means algorithm. The results showed the formation of three clusters with a Silhouette Score of 0.6213, indicating good clustering quality. The resulting clusters represent the categories of non-urgent, semi-urgent, and complex care needs patients. These segmentation results can support more systematic service prioritization, improve efficiency, accelerate response times, and support a data-driven triage system at the Langsa Regional General Hospital Emergency Department in Aceh.*

Keywords: *K-Means Clustering, Medical Informatics, Patient Care Prioritization, Patient Segmentation, Healthcare Data Mining*

Abstrak: Pelayanan pada Instalasi Gawat Darurat (IGD) menuntut kecepatan, ketepatan, dan koordinasi antar tenaga medis. RSUD Langsa Aceh sebagai rumah sakit tipe B menghadapi tingginya jumlah kunjungan pasien yang berpotensi menimbulkan ketidakefisienan, khususnya dalam penentuan prioritas pelayanan. Penelitian ini bertujuan mengoptimalkan segmentasi pasien IGD untuk mendukung penentuan prioritas pelayanan menggunakan algoritma K-Means clustering. Data yang digunakan merupakan data pasien IGD RSUD Langsa tahun 2025 sebanyak 31.761 record, dan setelah preprocessing diperoleh 24.062 data yang memenuhi kriteria analisis. Variabel penelitian meliputi frekuensi kunjungan, interval kunjungan, jenis layanan, durasi pelayanan, dan tingkat urgensi pasien. Metode penelitian menggunakan pendekatan kuantitatif melalui teknik data mining dengan algoritma K-Means. Hasil penelitian menunjukkan terbentuk tiga cluster dengan nilai Silhouette Score sebesar 0,6213 yang mengindikasikan kualitas pengelompokan baik. Cluster yang dihasilkan merepresentasikan kategori pasien non-urgensi, semi-urgensi, dan kebutuhan pelayanan kompleks. Hasil segmentasi ini dapat mendukung penentuan prioritas pelayanan secara lebih sistematis, meningkatkan efisiensi, mempercepat waktu respon, serta mendukung sistem triase berbasis data di IGD RSUD Langsa Aceh.

Kata Kunci: K-Means Clustering, Informatika Medis, Prioritas Pelayanan Pasien, Segmentasi Pasien, Data Mining Kesehatan

INTRODUCTION

Emergency Department (ED) care is one of the major challenges in healthcare, requiring a rapid and appropriate response. The ED is a highly complex care unit because it must treat patients with varying levels of urgency within a limited timeframe. (Di et al., 2025). According to data from the Aceh Provincial Central Statistics Agency (2026) in the 2025 Aceh Provincial Health Statistics, Type B hospitals in Aceh handle an average of 80,000–100,000 Emergency Department (ED) visits per year. Langsa Aceh Regional General Hospital, as one of the Type B hospitals in eastern Aceh, faces similar challenges in managing patient care. The high number of visits has the potential to cause problems such as long queues, delays in patient treatment, and suboptimal use of data, all of which affect the quality of ED patient care.

Although hospital management information systems (HMIS) have been able to provide comprehensive patient data, the use of this data is generally still limited to the presentation of descriptive information. Stored medical records of emergency department patients have not yet been fully processed to generate pattern-based information that can be used to support more effective clinical decision-making.

In this context, medical informatics through a data mining approach can be utilized to process hospital operational data into more meaningful information. One method that can be used is the K-Means clustering algorithm, which is capable of grouping data based on similar characteristics to produce relevant patient segmentation (Han et al., 2012).

Previous research on clustering in the health sector has generally focused on clinical aspects such as disease diagnosis. However, research utilizing hospital operational data to support the optimization of patient service priorities remains limited, particularly in Type B government hospitals in Aceh. Therefore, an approach is needed that can integrate data analysis with the operational needs of healthcare services.

Given these challenges, this study aims to apply the K-Means clustering algorithm to segment patients based on medical record data from Langsa Regional General Hospital in Aceh for the year 2025. The segmentation results are expected to provide additional information to support patient prioritization and enhance the effectiveness of interprofessional collaboration in healthcare delivery. This study focuses on the use of data-driven patient triage.

METHOD

This study employs a quantitative approach based on data mining, utilizing the K-Means clustering algorithm to group medical record data from patients at Langsa Regional General Hospital in Aceh for the year 2025. It also supports the concept of data-driven decision-making within the triage system, where prioritization decisions are not solely based on intuition but are also supported by historical data patterns.

The objective of this clustering is to generate patient segmentation based on data patterns that can support the prioritization process for emergency department (ED) patient care, thereby improving overall service quality. The K-Means algorithm is a non-hierarchical clustering method used to group data into a number of clusters based on the proximity of the data points. This algorithm works by minimizing the distance between the data and the cluster center (centroid), so that data within a single cluster share similar characteristics, while there are significant differences between clusters.

The data used in this study consist of:

1. Primary data, namely medical records of patients at the Emergency

Department (ED) of Langsa Regional General Hospital in 2025, totaling 31,761 records obtained from the Hospital Management Information System (HMIS). After preprocessing, the number of data points used in the analysis was 24,062.
2. Secondary data, namely health service statistics sourced from the official publication of the 2025 Aceh Province Health Statistics (Aceh Province Central Statistics Agency, 2026). Secondary data was used as a basis for describing the context of patient visit numbers and the hospital's service workload. Estimates of ED patient visit numbers were compiled based on average data for Type B hospitals in Aceh and the service capacity of Langsa Regional General Hospital, such as the number of active clinics, days of service, and the average number of patients per day.

The variables in this study consist of input variables (clustering) and output variables (evaluation indicators):

Input Variables (K-Means Clustering)

The service type variable was not used in the clustering process because all data were derived from the ED.

The study covered five operational patient dimensions:

1. Visit frequency (times/year, range 0–25)
2. Interval between visits (days, range 0.1–180)
3. Service duration (minutes, range 5–240)
4. Urgency level (1 = high, 3 = low)

Output Variables (Patient Care Priority Indicators)

Used to evaluate the impact of the clustering results:

1. Service response time
2. Service efficiency
3. Priority scale and evaluation of medical staff performance
4. Accuracy of patient prioritization

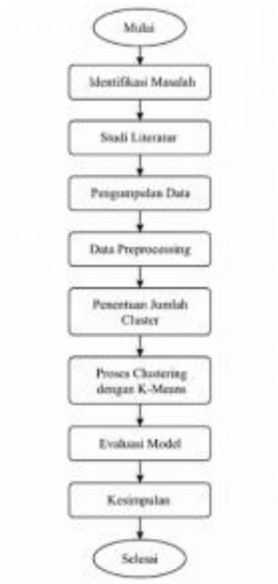

Figure 1 Research Flow

Figure 1 shows the research workflow, which begins with a literature review to establish a theoretical foundation regarding patient clustering based on patient priority levels and machine learning, followed by data collection, preprocessing to clean the data, and model evaluation. The following are the steps carried out in this study:

Preprocessing Data

The initial stage is conducted to ensure data quality, including:

1. Data cleaning by removing duplicate data
2. Handling missing values (<5%)
3. Transforming categorical data using label encoding
4. Data normalization using the Min-Max normalization method with a scale of 0–1

Determining the Number of Clusters

The optimal number of clusters is determined using the Elbow Method based on the Within-Cluster Sum of Squares (WCSS) value. The analysis results indicate an optimal value at K=3

Clustering Process with K-Means

The K-Means algorithm is used to cluster patient data by minimizing the

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
April 2026, IX (2): 2581 – 2587 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

distance between the data and the centroid in each cluster. The following is the K-Means formula

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - c_i||^2$$

Component Descriptions:

1. J = Total variation within clusters (inertia).
2. k = Number of clusters
3. Ci = The i-th cluster.
4. x = A data point that is a member of cluster Ci
5. µi = The centroid of cluster Ci
6. $||x - \mu i||^2$ = The Euclidean distance squared between data point x and cluster centroid µi

Parameters used:

1. n_init = 10
2. max_iter = 300
3. random_state = 42

The objective function in the K-Means algorithm aims to minimize the total distance between each data point and the centroid of its respective cluster. The value J represents the sum of squared distances between the data points and the cluster centers; thus, the smaller the value of J, the better the quality of the data clustering. This process is performed iteratively, where data is continuously grouped into the nearest centroid, and the centroids are updated until no significant changes occur or the specified iteration limit is reached.

The parameter n_init = 10 indicates that the centroid initialization process is performed 10 times to obtain the best results; max_iter = 300 is the maximum iteration limit in the clustering process; and random_state = 42 is used to ensure that the clustering results are consistent and reproducible.

Model Evaluation

Evaluation was performed using the Silhouette Score to measure the quality of separation between clusters. The evaluation results show a value of 0.6213, indicating that the clustering quality is in the good category.

Impact Analysis

The clustering results were analyzed using a pre-post simulation approach on indicators of patient service priority levels, including:

1. Medical staff response time
2. Service efficiency
3. Accuracy of patient priority

RESULT AND DISCUSSION

Data Preprocessing Results

The initial dataset consisted of 31,761 records obtained from the extraction of medical records of emergency department patients at Langsa Regional General Hospital in 2025. After preprocessing—which included removing missing values and filtering outliers in the service duration variable (0 < duration < 1,440 minutes)—the dataset used in the analysis was reduced to 24,062 records. This reduction aimed to improve data quality, ensuring that the resulting clustering analysis was more accurate and representative.

Clustering Results

The application of the K-Means algorithm with three clusters (k=3) yielded a Silhouette Score of 0.6213, indicating that the quality of cluster separation falls into the "good" category. This suggests that each cluster possesses distinct characteristics and is well-separated.

Visualization of Clustering Results

The visualization of the clustering results is presented in the form of a scatter plot illustrating the relationship between visit frequency and patient service duration.

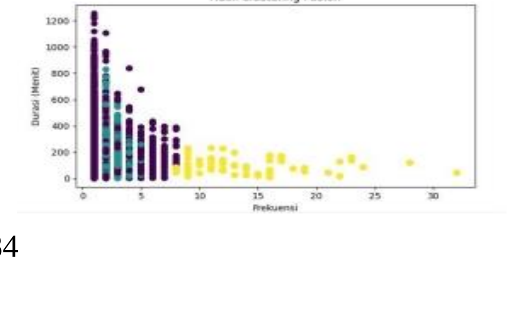

Figure 2 Visualization of patient clustering results using the K-Means algorithm (Scatter plot of visit frequency and service duration by cluster)

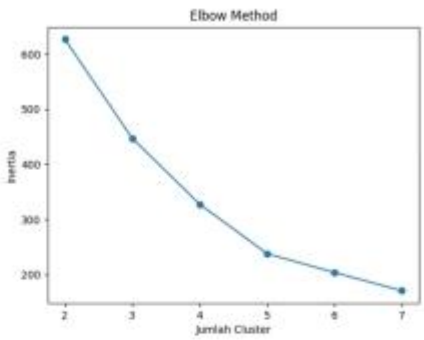

Figure 3 Visualization of the Results of the Elbow Method

Patient Segmentation System Architecture

The patient segmentation model developed in this study can be integrated with the Hospital Management Information System (HMIS) as a decision support system.

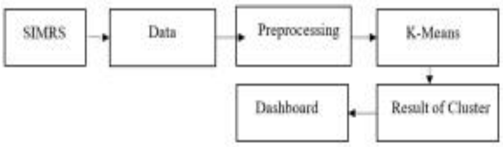

Figure 4 Segmentation System Flow

Figure 4 shows the segmentation system workflow, which includes data extraction from the SIMRS, data preprocessing, and clustering using the K-Means algorithm.

Cluster Characteristics

Based on the clustering results, the average characteristics of each cluster are as follows:

Cluster	Frequency	Interval (day)	Type of Services	Duration (minute)	Urgency
0	14,90	13,53	1,00	640,26	3,00
1	5,62	4,35	1,00	81,37	3,00
2	14,77	5,50	1,00	360,73	2,99

Table 1 Cluster Characteristics

Table 1 shows the results of clustering three groups of patient characteristics with good separation quality. These three clusters represent the categories of non-urgent patients, semi-urgent patients, and patients with complex care needs.

Cluster Interpretation

Based on the characteristics obtained, each cluster can be interpreted as follows:

1. Cluster 1 - Non-Urgent Patients (Fast Track)
   This cluster has a relatively low visit frequency with the shortest service duration; it can be said that these patients can be handled through a fast-track service flow without high priority. This indicates that patients in this group generally have mild conditions and do not require complex management.
2. Cluster 2 - Semi-Urgent Patients
   This cluster shows a fairly high visit frequency with moderate service duration; it can be said that these patients require observation and moderate priority. This group represents patients with moderate care needs who require more attention than patients with mild conditions.
3. Cluster 0 - Complex Patients (High Care Priority)
   This cluster has the longest service duration and a high visit frequency with longer intervals, indicating a need for high priority in care. This suggests that patients in this group have more complex conditions and require a longer treatment time.

Impact on the Quality of Emergency Department Patient Care

Based on the results of patient segmentation using the K-Means algorithm, there is potential for improved effectiveness in prioritizing emergency department patient care. The resulting segmentation allows medical staff to group patients based on urgency level and care needs, thereby supporting a more

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
April 2026, IX (2): 2581 – 2587 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

systematic and targeted prioritization process.

By grouping patients into non-urgent, semi-urgent, and complex categories, healthcare providers can:

1. Optimize service time allocation
2. Improve accuracy in determining patient priorities
3. Establish priority scales and evaluate healthcare staff performance

This demonstrates that a data mining-based approach has the potential to enhance service efficiency and the responsiveness of healthcare staff in managing patients.

Discussion

The results of this study indicate that although the hospital information system provides comprehensive data, its utilization is still limited to descriptive data presentation. Through the application of K-Means clustering, this data can be processed into pattern-based information that adds value to the decision-making process.

The resulting segmentation enables:

1. More objective identification of priority patients
2. Reduction of the cognitive load on medical staff during decision-making
2. Improvement in the efficiency and speed of patient care

In the context of overcrowding, this approach can assist in managing patient flow more efficiently. This aligns with the concept of data-driven decision-making in medical informatics, where decisions are supported by historical data analysis.

However, this study has limitations, including:

1. The analysis process is conducted offline and is not yet directly integrated with the SIMRS system
2. The service type variable does not show significant variation
2. Clustering results still depend on the quality of data preprocessing

Therefore, future research is recommended to develop real-time system integration and incorporate a wider variety of variables to improve model accuracy.

CONCLUSION

Based on the results of this study, it can be concluded that the application of the K-Means clustering algorithm to medical record data of emergency department patients at Langsa Regional General Hospital in Aceh was able to produce meaningful patient segmentation based on operational service characteristics.

The clustering process applied to 24,062 data points that had undergone preprocessing resulted in three patient groups with good separation quality, as indicated by a Silhouette Score of 0.6213. These three clusters represent the categories of non-urgent patients, semi-urgent patients, and patients with complex service needs.

The segmentation results indicate that the variables of visit frequency, visit interval, and service duration play a significant role in distinguishing patient characteristics. This segmentation provides additional information not immediately apparent in the raw data, thereby serving as a foundation to support healthcare decision-making. Furthermore, the results of this study indicate that a data mining-based approach can support the optimization of patient prioritization in the ED by identifying patients' levels of need and urgency in a more systematic and data-driven manner.

REFERENCES

Di, P. *et al.* (2025) "Pengaruh Pelayanan Kegawatdaruratan Terhadap Response Time Perawat Di Igd Di Rsud Toto Kabila Kabupaten Bone Bolango," 10(2), pp. 27–39.

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
April 2026, IX (2): 2581 – 2587 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

World Health Organization. (2010). Framework for action on interprofessional education and collaborative practice. WHO Press.

Badan Pusat Statistik Provinsi Aceh. (2026). Statistik kesehatan Provinsi Aceh 2025. BPS Provinsi Aceh. https://aceh.bps.go.id

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.

Tan, P. N., Steinbach, M., & Kumar, V. (2019). Introduction to data mining (2nd ed.). Pearson

Nugroho, Y. (2021). Analisis clustering K-Means pada data kesehatan. Jurnal Teknologi Informasi, 8(1), 12-20.

Sari, R., & Putra, A. (2022). Penerapan data mining dalam segmentasi pasien rumah sakit. Jurnal Informatika Kesehatan, 10(2), 45-52.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on Neural Networks, 16(3), 645–678.

Prasetyo, E. (2012). Data mining: Konsep dan aplikasi menggunakan MATLAB. Andi.