Bisantara Informatics Journal (JBI)
Vol.10, No.2, December 2025, pp. 1~10
ISSN (print): 2686-6455, ISSN (online): 2686-5319

Analysis of Patient Medical Records Using the K-Means Clustering Algorithm Based on Visit Time as a Service Strategy Approach

Eva Mufida Padilla [1], Aidul Safii [2], Indra Nasution [3], Maisya Fitri Anugrah [4], Nurlina sari Harahap[5], Muhammad Syahputra Novelan [6]
[12345] Postgraduate, Master of Information Technology, Universitas Pembangunan Panca Budi, Medan, Indonesia
email: [1] evamufidapadilla@gmail.com, [2] aidulsafii0707@gmail.com,
[3] sutanindrabungsu@gmail.com, [4] maisyafitrianugrah@gmail.com,
[5] putranovelan@dosen.pancabudi.ac.id

Abstract

This study develops an analytical framework to optimize service strategies in primary healthcare, focusing on midwifery clinics that use electronic medical records. It employs the K-Means clustering algorithm to segment patients by visit time, diagnosis, and demographic characteristics, addressing the limitations of intuition-based decision-making for fluctuating patient volumes and resource needs. The clustering results provide an objective basis for designing adaptive interventions in staffing schedules, queue management, and pharmaceutical inventory, with the overall aim of improving patient satisfaction and operational efficiency
***Keywords**: Medical Records; Patients; Visit Time; K-Means Clustering Algorithm.*

1. INTRODUCTION

In the modern healthcare service ecosystem, medical records are not merely archives of disease history, but strategic assets that contain patterns of patient behavior(Asminanda et al., 2025; Grant et al., 2020; Novaliendry et al., 2023; Ravuvar et al., 2020). However, these data often accumulate only as “digital waste” or physical documents without further processing(Prayogi et al., 2025). Klinik Bidan Neni, as a representation of a primary healthcare facility in an urban area such as Medan City, faces multidimensional challenges: high variability in patient arrival times and a diversity of service types (ranging from routine immunizations to emergency delivery care). The main problem is the misalignment between the availability of resources (midwives, medicines, waiting rooms) and the actual service demand. Service targets and schedules are often determined reactively and subjectively, resulting in long queues at certain hours and idle time at others(Haraty et al., 2015).

An objective method is needed to map “when” patients come and “what” they need. This analysis aims to transform raw medical record data into strategic insights using Data Mining methods(Alasi, 2024). Specifically, it seeks to identify patient(Sembiring et al., 2023) segments based on the correlation between Visit Time and Type of Disease/Service, and to design precise Service Strategies based on the characteristics of each resulting segment (cluster). Advances in information technology have brought major changes in many sectors, including healthcare. One form of its application is a medical record data management system used to record and store patient information digitally. Medical records not only function as documentation of a patient’s health history, but can also serve as a valuable data source for analysis and for formulating more effective healthcare service strategies(Hillerman et al., 2015). Klinik Bidan Neni in Medan City, as one of the healthcare facilities providing midwifery and maternal–child health services, receives many patients each day with varying visit times. The stored patient medical record data have not yet been fully utilized to analyze patterns of patient visits, even though such patterns can serve as a basis for decision-making in managing service

Received November 1st, 2025; Revised November 25th, 2025; Accepted Desember 1th, 2025

schedules(Lai et al., 2025; Ramadhan & Astuti, 2024), medical human resources, and improving service quality for the community. Data mining methods, particularly the K-Means Clustering algorithm(Hidayah et al., 2025), can be used to group patient data based on certain characteristics, including visit time. By applying this algorithm to patient medical record data, the clinic can identify the busiest visit time groups, patient arrival patterns, and potential improvements in service strategies according to needs. This study aims to analyze patient medical record data at Klinik Bidan Neni in Medan City using the K-Means Clustering algorithm based on visit time. The results of this study are expected to provide input for the clinic in developing service strategies that are more optimal, efficient, and data-driven, in order to support the improvement of the quality of public health services.

2. RESEARCH METHODS

The research process begins with collecting medical record data that includes time, clinical, and demographic attributes, followed by a literature review to establish a foundation for understanding the K-Means algorithm and clustering techniques. The data are then preprocessed through time conversion into numerical format, one-hot encoding for diagnosis and insurance variables, and normalization to ensure balanced feature scales. Next, the number of clusters is determined using the Elbow Method before the K-Means algorithm is executed to calculate Euclidean distance and assign each data point to the nearest centroid. The resulting clusters are evaluated based on intra- and inter-cluster variation, then interpreted to identify patient visit patterns such as morning, afternoon, or night. The final stage involves formulating service strategy recommendations based on these patterns, marking the completion of the research process.

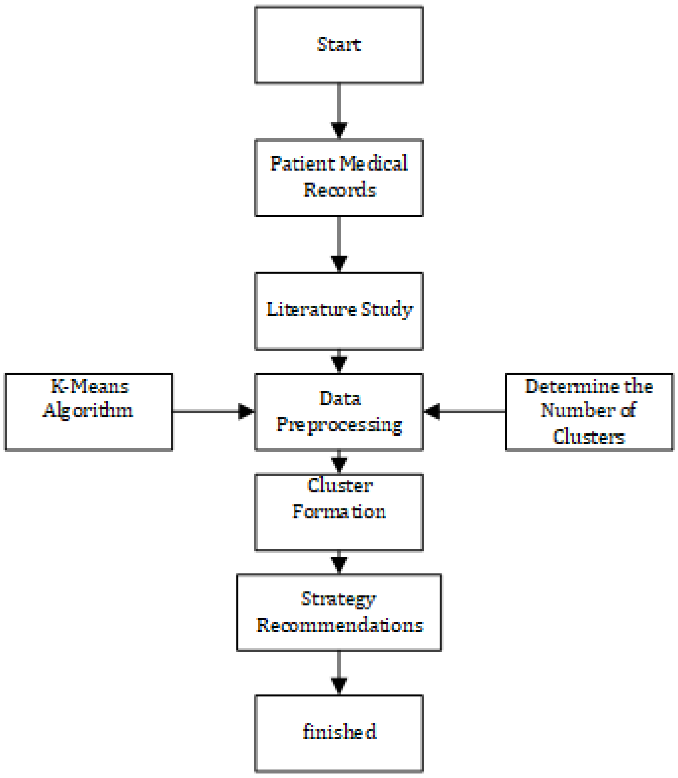

Fig 1. Research flow of Analysis of Patient Medical Records Using the K-Means Clustering

2.1 K-Means

K-Means Clustering is a method in which K represents the constant number of desired clusters, and "means" refers to the average value of a group of data, which in this context is defined as a cluster. Thus, K-means clustering is a data analysis or data mining technique that performs unsupervised modeling and uses a partition-based approach to group data. The K-means algorithm is a cluster analysis method that partitions N observations into clusters such that each observation belongs to the group with the closest mean. The K-means algorithm is simple to implement(Alasi, 2024; Nofriansyah, 2015; Prayogi et al., 2025), relatively fast, easy to adapt, and widely used. In K-means, data are classified based on their proximity to the centroid; the goal is to maximize similarity of data within the same cluster and minimize similarity of data between clusters(Yahya & Kurniawan, 2025).

Clustering is a technique for grouping data that share similar characteristics with one another. It is an unsupervised data mining method, applied without training data, without a teacher, and without requiring target outputs. In general, clustering methods for grouping data can be divided into two types: hierarchical clustering and non-hierarchical clustering.

This algorithm partitions patient data into clusters by minimizing variation within clusters (intra-cluster variation) and maximizing variation between clusters. Determining the Number of Clusters: The Elbow Method is used to find the optimal value of K, where adding more clusters no longer yields a significant reduction in variance. In general, for a primary clinic, 3–4 main clusters are formed based on time patterns (Morning, Afternoon, Night). Euclidean Distance: Used to measure similarity between patients. Patients who come at the same time with similar complaints will have a small Euclidean distance and will be grouped together

2.2 Patient Medical Records

The data analyzed are sourced from electronic medical records as well as manual registrations and include the following attributes: Time Dimension: Time of arrival, Day of visit. Clinical Dimension: Diagnosis (ICD-10 or descriptive), Vital signs (Blood Pressure, Temperature), Medical procedures. Demographic Dimension: Age, Address (District), Type of insurance (BPJS/General)

2.4 Data Preprocessing

To enable the algorithm to read visit patterns, the data are transformed as follows: Conversion of Time to Numeric: The visit time (e.g., '08:30') is converted into a decimal format or time bins (Morning, Afternoon, Evening) to facilitate Euclidean distance calculation. Encoding of Diagnosis Variables: One-Hot Encoding is used for nominal data such as disease type and insurance type, in order to prevent irrelevant ordinal bias (for example, treating disease code A as 'smaller' than disease code B). Normalization (Min-Max Scaling): The scales of age data (0–60 years) and time data (0–24 hours) are equalized so that no single variable dominates cluster formation.

2.5. Research Stages

The research methods used by the researcher to solve the problem at Klinik Bidan Neni in Medan City, located at Jl. Sampul Gg. Pribadi No. 35 A, Sei Putih Barat, Medan Petisah District, Medan City, North Sumatra, are as follows.

a. Literature Study

Obtaining information by collecting, studying, and reading various references from books and journals related to this research, such as intelligent systems, data mining, the k-means clustering algorithm, and others.

b. Field Study

Obtaining information by collecting and understanding data directly from Klinik Bidan Neni in Medan City.

1. Observation
   Conducting observations, directly observing the work procedures of the units related to recording the results of activities carried out, in order to see the working system of data processing in the process of patient visits and diseases.
2. Sampling
   The sampling technique is the selection of a certain number of items from all available items with the aim of studying part of those items to represent the entire system. One such sampling activity is carried out at Klinik Bidan Neni in Medan City.

3. RESULTS AND DISCUSSION

3.1 Data

Daily medical record data above reflect the diversity of health services provided in several areas of Medan. Each entry contains the patient's time of arrival, age, sex, address, insurance status, vital sign data such as blood pressure, body temperature, body weight, main diagnosis, ICD-10 code, and types of services ranging from Outpatient, Midwifery, Immunization, Family Planning, Prolanis, Counseling, Emergency Department, to Delivery Room/Maternity. The variation in diagnoses, ranging from common infectious diseases and chronic conditions to pregnancy, immunization, emergency cases, and obstetric care, illustrates the scope and intensity of integrated health services. These data are processed in a structured format, making it easier to perform clinical analysis, service evaluation, and public health monitoring.

Tabel 1. Source Data Process

No	In	Old	G	address	By	Vital_TD	temp	BB	Weight	Room	Type
1	07:15	4	L	Medan Tuntungan	Umum	-	38.5	14	Febris / Fever	R50.9	Outpatient
2	07:30	2	P	Medan Johor	BPJS	-	37.8	11	Acute Respiratory Infection (ARI)	J06.9	Outpatient
3	07:45	28	P	Medan Selayang	Umum	110/70	36.5	55	Hyperemesis Gravidarum	O21.0	Midwifery
4	08:00	5	L	Medan Tuntungan	BPJS	-	36.8	18	Diare Akut	A09	Outpatient
5	08:10	1	P	Medan Johor	Umum	-	36.6	9	Measles Vaccination	Z24.4	Imunisasi
6	08:2	32	P	Medan	BPJS	120/80	36.	62	Check-up	Z30.	KB

	5			Amplas			7		For IUD	5	
...	...	...	...	...	...	...	...	...	...	...	...
100	09:00	0	L	Medan Johor	BPJS	-	36.9	4	Neonatal Check-up	Z38	Midwifery

Tabel 2. *Pre-Processed*

N0	in	old	servises	Urgenci	BPJS_Status	Blood Pressure	Time
1	7.25	0.07	1	4	0	0	1
2	7.50	0.03	1	3	1	0	1
3	7.75	0.47	2	3	0	110	1
44	8.00	0.08	1	3	1	0	1
5	8.17	0.02	3	1	0	0	1
6	8.42	0.53	4	1	1	120	1
7	8.67	0.10	1	4	0	0	1
8	9.00	0.00	2	2	1	0	1
...	...	...	...	...	...	...	...
100	9.50	0.42	1	2	1	90	1

This graph visualizes patient groupings based on two main variables: Hour of Visit (X-Axis) and Patient Age (Y-Axis). The algorithm has divided the data into 3 different clusters, marked with red, turquoise and blue.

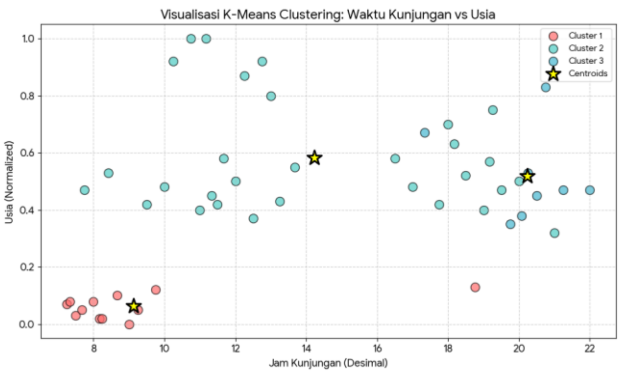

Figure 2. Patient Medical Record Results Based on Visit Time in Service Strategy Efforts.

This graph visualizes patient clustering based on two key variables: Visiting Hour on the X-axis and Patient Age on the Y-axis. The algorithm has separated the data into three distinct clusters, indicated by the colors red, turquoise, and blue. Each circle on the graph represents a patient, and the color of the circle indicates the cluster resulting from the K-Means clustering, influenced by characteristics such as arrival time, age, and services received. Furthermore, the large yellow star represents the centroid, or center point, of each cluster; the K-Means algorithm iteratively shifts this centroid until an optimal position is reached, allowing patient groups to be visually identified based on their data similarities.

Each column of numeric data is designed using a specific transformation to ensure optimal algorithm readability. The Hours_Decimal column, for example, is generated by converting the hour format from a colon (e.g., 07:15) to a decimal fraction using the formula "Hours" + "Minutes" / 60, so that 07:15 becomes 7:25 and 07:30 becomes 7.5. The Service_Code column applies an encoding label by assigning a numeric code to the service category, such as Outpatient as 1, Midwifery as 2, Immunization as 3, etc., so that the computer can easily distinguish the types of services. The BPJS_Status column uses a binary encoding technique, namely 1 if the insurance is "BPJS" and 0 for "General". Systolic_Blood_Pressure is processed from raw vital sign data: if there is a number such as "120/80", the first digit (systole) is taken, while if it is empty ("-") it is filled with a value of 0 for consistency and to prevent errors in the analysis. Time_Category uses a binning technique to group patient visit times, for example, 7:00–11:00 am is given a code of 1, 11:01–15:00 pm is coded 2, and afternoon/evening is coded 3. Finally, Age_Norm is the result of normalizing the patient's age using the formula ("Age" -0)/(60-0) so that the age range can be compared proportionally with other variables. This processing ensures the data is neat, uniform, and easily processed by analysis algorithms.

This data has been numerically processed so that each variable represents important clinical information in a standardized format: the Decimal_Hour column indicates the patient's arrival time in decimal units, Age_Norm is the age-normalized result, Service_Code encodes the type of healthcare service, Urgency_Score represents the urgency of the case, BPJS_Status indicates the patient's insurance status (0 for general, 1 for BPJS), Systolic_Blood_Pressure records the systolic blood pressure value, and Time_Category helps group the data by specific time range.

Service category codes use numbers 1 through 8 to facilitate data management: 1 represents General Outpatient, 2 for Obstetrics (ANC/Gynecology), 3 for Immunization, 4 for Family Planning, 5 for Prolanis (Chronic), 6 for Counseling, 7 for Emergency Room, and 8 for VK/Maternity. The use of numeric codes aims to ensure efficient and uniform identification of service types within the information system.

Each attribute in the data is processed using appropriate transformation techniques to enable numerical analysis: the raw time "07:15" is converted using time conversion techniques to a value of 7.25; the age "4" is normalized using min-max normalization to a value of 0.07; the service "Outpatient" is converted to a value of 1 through label encoding; the urgency scale "4" is used directly because it is already numeric; the coverage "General" is encoded using binary encoding to a value of 0; missing vital signs are imputed and parsed to a value of 0; and missing categories are binned to a value of 1. These transformations ensure that all data is formatted uniformly and ready for further analysis.

Based on simulation results from applying the K-Means Clustering algorithm to historical patient medical records at the Neni Midwife Clinic in Medan, three main clusters have been clearly identified that directly impact the clinic's service strategy.

Cluster 1- Morning Acute Patients, During the analysis period, patients belonging to the first cluster were recorded as predominantly visiting the clinic between 7:00 AM and 10:00 AM WIB. Data shows that this group is dominated by toddlers and adults with acute complaints that appear suddenly, such as fever, acute respiratory infections (ARI), and diarrhea. The surge in visits during these hours is due to the tendency of parents to bring sick children the night before, right after the clinic opens in the morning. This category of patients is noted to have a low tolerance for waiting times due to their physical discomfort and need for immediate treatment. Cluster 2 - Chronic Patients & Midday Routine Cluster, Over time, the next visit pattern was detected during the day, namely between 11:00 AM and 2:00 PM WIB. The second cluster in this clustering analysis consists of elderly patients, pregnant women (first and second trimester ANC), and Prolanis program participants (hypertension and diabetes). Historical data shows that this group regularly comes for scheduled consultations and tends to avoid the morning crowds. Their consultations require relatively longer timeframes, particularly for education and counseling, but their medical conditions are generally stable, so services are preventive and monitoring. Cluster 3 - After-Hours & Emergency Patients, The third cluster was found to be most active in the afternoon and evening, between 5:00 PM and 9:00 PM WIB. Based on data analysis, this segment includes patients of productive age, including workers, family planning patients, and obstetric emergencies such as postpartum or hemorrhage. Data shows that workers can only visit after office hours, and the pattern of visits for childbirth or emergency cases often occurs randomly but peaks in intensity at night. This situation requires staff preparedness to provide immediate medical treatment. Formulating a Data-Driven Service Strateg. The clustering findings serve as the basis for developing a measurable and efficient data-driven service strategy at the clinic.

Based on patient load patterns from the analysis, midwife and nurse schedules have been dynamically adjusted to meet cluster needs. The morning shift is allocated with the largest staff, focusing on speedy administration and handling of symptomatic complaints to reduce queues in Cluster 1. The afternoon shift deploys senior midwives with strong communication skills, prioritizing quality consultation time, especially for education on chronic diseases and pregnancy from Cluster 2. The evening shift ensures the availability of certified APN midwives to ensure readiness for emergencies that frequently arise in Cluster 3. The distribution of medication and medical equipment has been adapted to suit the visit patterns of each cluster. Pediatric medications, such as syrup and paracetamol, as well as basic antibiotics, are routinely ensured to be fully stocked before morning opening hours, meeting the fast-moving needs of Cluster 1. Meanwhile, critical medications for emergency procedures, such as oxytocin and magnesium sulfate, as well as contraceptives, are always rechecked during the afternoon shift handover to ensure they are ready for use in the cases predominantly occurring in Cluster 3. With this analysis-based strategy, services at the Neni Midwife Clinic in Medan City have been directed to be more precise and effective, based on the characteristics of patient visit times from historical medical record data.

4. CONCLUSION

The application of the K-Means Clustering algorithm to the medical record data of Neni Midwife Clinic shows that visit time is closely linked to disease patterns and patient profiles, enabling a shift from intuition-based decisions to a fully data-driven service model. By analyzing historical records, three dominant clusters emerge: acute morning patients, chronic and follow-up midday patients, and after-hours workers combined with obstetric emergencies. The morning cluster is dominated by children and adults with sudden-onset complaints such as fever, respiratory infections, and diarrhea, requiring fast service and minimal waiting time. The midday cluster consists largely of elderly patients, pregnant women in routine antenatal care, and Prolanis participants with chronic conditions, who need longer consultation time for education and monitoring rather than emergency management. The evening cluster is characterized by productive-age workers, family planning clients, and emergency obstetric cases that demand readiness for immediate intervention. Using these clusters, the clinic can calibrate staffing across shifts, align pharmaceutical inventory with time-based demand, and reorganize service flow according to patients' biological and social rhythms. As a result, operational efficiency, waiting-time management, and the overall quality of midwifery and primary care for the surrounding community are significantly

.

5. SUGGESTIONS

Future research should integrate user-centered design principles into the K-Means clustering model by incorporating real-time patient feedback through mobile applications or clinic portals, allowing dynamic adjustment of clusters based on evolving user profiles and preferences. This integrated approach would address current limitations in static historical data analysis, enabling personalized scheduling recommendations that align staffing and resources more precisely with patient-reported needs and behaviors. Such enhancements could validate and refine the identified clusters—acute morning, chronic midday, and evening emergency groups—via longitudinal user studies, ultimately improving operational efficiency and care quality in midwifery settings.

THANK-YOU NOTE

We would like to express my sincere gratitude to the Jurnal Bisantara Informatika for accepting and publishing this article. Our appreciation also goes to all readers who have taken the time to read and engage with this work. We hope this article brings valuable insights and contributes positively to the field of education, especially in computer science. Any suggestions and feedback are greatly appreciated for the improvement of future research.

BIBLIOGRAPHY

	Eva Mufida Padilla, Currently pursuing a Master's degree in Computer Science with a specialization in Information Technology at Universitas Pembangunan Panca Budi.
	Aidul Safii, Currently pursuing a Master's degree in Computer Science with a specialization in Information Technology at Universitas Pembangunan Panca Budi.

	Indra Nasution, Currently pursuing a Master's degree in Computer Science with a specialization in Information Technology at Universitas Pembangunan Panca Budi.
	Maisya Fitri Anugrah, Currently pursuing a Master's degree in Computer Science with a specialization in Information Technology at Universitas Pembangunan Panca Budi.
	Nurlina sari Harahap, Currently pursuing a Master's degree in Computer Science with a specialization in Information Technology at Universitas Pembangunan Panca Budi.
	Muhammad Syahputra Novelan, Bachelor, Master, and Doctor of Computer Science; Intelligent Systems Lecturer at Universitas Pembangunan Panca Budi. Holds the academic rank of Senior Lecturer (Associate Professor in the field of Machine Learning)

REFERENCES

Alasi, T. S. (2024). *Ilmu Komputer* (M. Ihsan (ed.); 1st ed.). https://www.media-publikasi-idpress.my.id/2023/12/ilmu-komputer.html

Asminanda, C. R., Nisa, W., Idris, I., Sari, R. A., & Khair, R. (2025). Optimization of hospital patient management through a combined RFM and fuzzy C-Means approach: enhancing service efficiency and quality. *Journal of Science and Education (JSE)*, *5*(2), 458–466.

Grant, R. W., McCloskey, J., Hatfield, M., Uratsu, C., Ralston, J. D., Bayliss, E., & Kennedy, C. J. (2020). Use of latent class analysis and k-means clustering to identify complex patient profiles. *JAMA Network Open*, *3*(12), e2029068--e2029068.

Haraty, R. A., Dimishkieh, M., & Masud, M. (2015). An enhanced k-means clustering algorithm for pattern discovery in healthcare data. *International Journal of Distributed Sensor Networks, 11*(6), 615740.

Hidayah, A. H., Novitasari, D., Kamila, R., Nasrudin, M., & others. (2025). Cluster Modeling with K-Means on Provincial Data in Indonesia Based on Environmental Indicators. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *4*(3), 2245–2249.

Hillerman, T. P., Carvalho, R. N., & Reis, A. C. B. (2015). Analyzing suspicious medical visit claims from individual healthcare service providers using k-means clustering. *International Conference on Electronic Government and the Information Systems Perspective*, 191–205.

Lai, H., Huang, T., Lu, B., Zhang, S., & Xiaog, R. (2025). Silhouette coefficient-based weighting k-means algorithm. *Neural Computing and Applications*, *37*(5), 3061–3075.

Nofriansyah, D. (2015). *Konsep Data Mining Vs Sistem Pendukung Keputusan*.

Novaliendry, D., Wibowo, T., Ardi, N., Evi, T., & Admojo, D. (2023). Optimizing Patient Medical Records Grouping through Data Mining and K-Means Clustering Algorithm: A Case Study at RSUD Mohammad Natsir Solok. *International Journal of Online \& Biomedical Engineering*, *19*(12).

Prayogi, S. Y., Alasi, T. S., & Rahmat, R. F. (2025). *Pengantar Machine Learning* (M. Ihsan (ed.); 1st ed.). Media Publikasi Idpress. https://www.media-publikasi-idpress.my.id/2025/03/4.html

Ramadhan, G., & Astuti, Y. (2024). Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail. *Jurnal Informatika: Jurnal Pengembangan IT*, *9*(1), 92–96.

Ravuvar, H. N., Goda, H., Chinnasamy, P., & others. (2020). Smart health predicting SystemUsing K-means algorithm. *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 1–4.

Sembiring, A. S., Alasi, T. S., & others. (2023). Penyedia Layanan Konsultasi Kesehatan dengan Metode TOPSIS. *Jurnal Armada Informatika*, *7*(1), 274–280.

Yahya, A., & Kurniawan, R. (2025). Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan Berdasarkan Pola Penjualan: Implementation of K-Means Algorithm for Clustering Sales Data Based on Sales Patterns. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(1), 350–358.