Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

K-Means Clustering Algorithm Analysis For Grouping Patient Medical Record Data Based On Disease Type

Wina Aulia[1], Andysah Putera Utama Siahaan[2], Leni Marlina[3], Khairul[4], Muhammad Iqbal[5]
[1,2,3,4,5]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi

Article Info
Keywords:
Data Mining,
Medical Records,
K-Means Clustering Algorithm,
Python-based Jupyter Notebook

ABSTRACT
Dr. Tengku Mansyur Regional Hospital, Tanjungbalai City, still has difficulty in grouping the intensity of diseases that often occur in Tanjungbalai City, in the exact grouping based on the results of patient medical data records. Medical records are documents that record information about the patient's condition, medical history, and previous treatment filled in by health workers who provide care and treatment. Medical records function to provide health information for all health workers involved in patient care. The goal is to minimize the increase in the intensity of the disease experienced by patients by providing counseling and solutions to diseases that have high intensity. administration in an effort to improve health services in hospitals and also increase the supply of drugs according to the drugs needed. This study uses the Jupyter Notebook application based on Python and data mining methods, especially the K-Means Clustering algorithm, to analyze patient medical record data based on age, blood pressure and disease diagnosis. With the K-Means Clustering Algorithm, it can also minimize variation in one cluster and maximize variation between clusters. With the research, Cluster C1 is a disease that often occurs with high intensity and has 49 members, Cluster C2 is a disease that occurs with intensity. Cluster 2 has 28 members, while Cluster C3 is a disease that occurs with low intensity and has 23 members.

Corresponding Author:
Wina Aulia
Magister Teknologi Informasi, Universitas Pembangunan Panca Budi
winaaulia397@gmail.com

INTRODUCTION

Data Mining[1], [2], [3], [4] is the process of analyzing and extracting useful patterns or information from large data sets using statistical techniques, mathematics, and computer-based algorithms. Data mining is often used to identify hidden relationships, trends, or specific patterns in data that can be used to support decision making in various fields, such as business, health, education, finance, and technology.

Hospitals play an important role as health care facilities that provide comprehensive services, such as inpatient, outpatient, and emergency care. As a public service institution, hospitals have a responsibility to fulfill basic human rights in obtaining health services. The main goal is to provide high-quality health services that can satisfy patients. Hospital services are not only limited to medical procedures, but also include supporting services, one of which is medical records.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Medical records are documents that record patient information, medical history, and previous treatments made by health workers. Medical records have an important role in ensuring effective communication between health workers involved in patient care. Indicators of the quality of medical record services include completeness, accuracy, and speed in providing information needed for health services [5]. In the health sector, medical records are often associated with the International Classification of Diseases (ICD), a standard coding system used to record patient diagnoses and treatments [6]. This code allows consistent communication between health workers worldwide. Medical records contain important information about "who, what, why, when, and how" services are provided to patients, and provide sufficient detail to identify patients, justify diagnoses, and record treatment outcomes. In addition, medical records include patient identity, examination records, treatments, and other health services (Minister of Health Regulation No. 55 of 2013). The main purpose of medical records is to support the administrative process to improve the quality of health services.

Along with the advancement of information technology, grouping patient medical record data based on disease type is becoming increasingly important. However, at Dr. Tengku Mansyur Hospital, Tanjungbalai City, there is currently no feature that can group diseases based on their type. This need emphasizes the importance of a systematic approach to grouping diseases in order to meet patient needs effectively. The purpose of this study is to group diseases based on the intensity experienced by patients at Dr. Tengku Mansyur Hospital, Tanjungbalai City, so that the hospital can provide targeted health education and ensure the availability of the necessary drugs.

This study uses the K-Means Clustering algorithm, one of the data mining techniques that is recognized as efficient in unsupervised data clustering. By minimizing variation within a group and maximizing variation between groups, the K-Means method can provide useful insights in grouping patient medical record data [7]. Clustering, as an important part of data mining, aims to organize data into groups based on their similar characteristics, so that patterns and trends in large-scale data can be revealed [8]..

Dr. Tengku Mansyur Regional Hospital, Tanjungbalai City, which was established in 1930 during the Dutch colonial period, is one of the oldest hospitals in North Sumatra. Located in Tanjungbalai City, this hospital is also a referral center for patients from surrounding areas, such as Asahan, Batubara, and Labuhan Batu Regencies. On average, this hospital serves around 100 patients per day, or more than 24,000 patients per year by 2023. With a large number of patients, the application of data mining techniques such as K-Means Clustering can significantly improve the hospital's ability to manage and utilize medical record data.

This study aims to apply the K-Means algorithm [9], [10], [11], [12] Clustering in grouping patient medical records at Dr. Tengku Mansyur Hospital, Tanjungbalai City based on disease type. The insights generated from this grouping process are expected to support hospitals in handling public health problems related to high-intensity diseases and optimizing drug supplies. By addressing these challenges, this study aims to contribute to improving the quality of health services at Dr. Tengku Mansyur Hospital, Tanjungbalai City.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

RESEARCH METHODOLOGY

This study uses a quantitative approach to analyze patient medical record data at Dr. Tengku Mansyur Regional Hospital, Tanjungbalai City. The research stages consist of the following steps:

1. Literature Study: The research began by collecting references and related literature, including journals, scientific articles, and documents relevant to the use of the K-Means Clustering algorithm in medical record data analysis.
2. Data and Information Collection: Data were obtained through interviews with the hospital, especially Mrs. Anne as the Head of the Patient Medical Record Room at Dr. Tengku Mansyur Regional Hospital, Tanjungbalai City. The data collected included patient medical records with a focus on the type of disease and the intensity of patient visits.
3. Data Analysis with K-Means Clustering: The collected data is analyzed using the K-Means Clustering algorithm to group the data based on certain similarities or characteristics. The goal is to identify the types of diseases that patients experience most often.
4. System Design: The data from the analysis is used to design a system that utilizes the K-Means Clustering algorithm. This system is designed using the Python-based Jupyter Notebook application.
5. System Testing: The system is tested using patient medical record data to ensure its accuracy in grouping disease types according to the applied algorithm.
6. Results and Implementation: At this stage, the study produced a classification of diseases based on the intensity of patient visits. These results are used to provide recommendations for drug management and health education to the community.

Means Clustering Method

In the medical context, clustering disease types based on patient Medical Record data plays a crucial role in improving the accuracy and speed of diagnosis. The K-Means Clustering algorithm has proven to be effective as a method for analyzing Medical Record Data. K-Means Clustering is a method used to group data into several parts based on distance, which can only work on numeric attributes. This method falls under the category of clustering partitioning, which separates data into k different regions. K-Means is a non-hierarchical method where cluster centers are randomly selected from the data set and each data is placed into the nearest cluster based on the minimum distance. The cluster center positions are updated iteratively until all data is classified into its clusters, creating new cluster center positions.

The following are the steps of the K-Means Clustering algorithm.*[14], [15], [16]:*

1. Determine the number of k-clusters you want to form.
2. Determine random values for the initial cluster centers (centroids) for k-clusters.
3. Calculate all data with centroid using the distance formula (Eucledian Distance) until the closest distance of each data to the centroid is found.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Research Data

Selecting a representative sample from the population is very important so that the conclusions generated from the study can be applied more generally to the entire population. By using medical record data from 100 patients, researchers can analyze disease patterns, response to treatment, or other characteristics of the patient population that can be useful for improving health services at Dr. Tengku Mansyur Hospital, Tanjungbalai City. Sample data from 100 patients are presented in table 1.

Table 1. Patient Data at Dr. Tengku Mansyur Regional Hospital, Tanjungbalai City

No	Patient Name	Age	Tension	Diagnosis
1	The Greatest	59 years old	135	Diabetes, COPD
2	Liang Hong	65 years old	150	Hypertension, Diabetes
3	Religious	56 years old	122	Diabetes
4	Fatimah Syam	71 years old	131	Diabetes
5	Efi Erikawani	55 years old	130	Asthma Bronchitis
6	The Great Wall	74 years old	140	Diabetes
7	Minah Samosir	46 years old	166	Diabetes
8	Husnah Sitorus	59 years old	145	Diabetes
..	..	..	..	..
..	...	..	..	..
99	Lasmini	58 years old	135	Diabetes & Kidney Disorders
100	Zuniar	63 years old	140	Diabetes & Hypertension

Operational Definition of Research Variables

Operational definition of variables is an important step in research design, as it helps to measure and observe concepts that may be abstract or complex. In the context of research on the application of data mining for grouping medical record data based on disease type inDr. Tengku Mansyur Regional Hospital, Tanjungbalai CityFollowing are operational definitions for several variables.The following is an example of an operational definition for several relevant variables in the context of applying data mining for classifying medical record data based on disease type at Dr. Tengku Mansyur Regional Hospital, Tanjungbalai City:

1. Independent Variables
   a. Patient Age: Operational definition: The patient's age at the time of the first visit to Dr. Tengku Mansyur Regional Hospital, Tanjungbalai City, calculated in full years. Age Table
   b. Patient Gender: Operational definition: Patient gender, identified as male or female based on data at the time of registration.
   c. Medical History: Operational definition: Complete information regarding the patient's medical history, including previous diagnoses, medical procedures, and other relevant health conditions, obtained from the electronic medical records of Dr. Tengku Mansyur Hospital, Tanjungbalai City.
2. Dependent Variable (Type of Disease):
   Operational definition: Classification of disease types based on the International

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Classification of Diseases (ICD) codes listed in the patient's medical record. For example, heart disease (ICD-10: I00-I99), diabetes (ICD-10: E10-E14), etc.

3. K-Means Clustering Method
   a. Number of Clusters (K): The number of groups or clusters to be formed. formed based on K-Means analysis.
   b. *Centroid*Cluster: The center point of each cluster generated by K-Means algorithm.
   c. Classification Criteria: Criteria used to group patient medical record data based on disease type, especially internal diseases, for example based on age, blood pressure and diagnosis.

RESULTS AND DISCUSSION

Data Processing

The process of K-Menas clustering method based on patient medical record data from Dr. Tengku Mansyur Hospital, Tanjungbalai City, then data that has nominal and non-nominal data types such as patient name, age, blood pressure and diagnosis, must be initialized first into numbers. Grouping of this medical record data can be expressed in independent variables such as age (X), blood pressure (Y), diagnosis (Z).

Table 2. Age

Code	Age	Age Category
1	19 – 25 years old	Teenager
2	26 – 45 years old	Mature
3	46 – 75 Years	Elderly

Source: (Ministry of Health Regulation No. 25 of 2016)

Table 3. Tension

Code	Tension
1	90 – 120
2	120 – 150
3	150 – 180
4	180 – 210

Source: (Ministry of Health HK.01.07 of 2021)

Table 4. Diagnosis

Code	Diagnosis
1	Diabetes
2	Hypertension
3	Lungs
4	Stroke
5	Muscle Pain
6	Coronary heart
7	Prostate
8	Vertigo

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

9	Joint pain
10	Vomiting blood
11	Low blood sugar
12	Stomach acid
13	Antibiotic Drug Allergy
14	Indigestion
15	Chronic intestinal inflammation
16	Toxic Goiter
17	Kidney Disorders
18	Back Pain
19	Asthma Bronchitis
20	Chronic Heart Failure
21	Diabetes, COPD
22	Diabetes & Hypertension
23	Hypertension & Cholesterol
24	Diabetes & Kidney Disorders
25	Digestive Disorders & Vertigo
26	Chronic Heart Failure & Hypertension
27	Pulmonary Tuberculosis & Chronic Kidney Failure
28	Gastric dyspepsia & Cholesterol
29	Postpartum hemorrhage

Source: (RSU Dr. Tengku Mansyur 2023)

Input Variable Data

The input variable data to be inputted is patient data generated in the pre-process cleaning and transformation stages that were previously carried out. The stages are carried out using Microsoft Excel software. The dataset used is patient medical record data at Dr. Tengku Mansyur Hospital, Tanjungbalai City, the patient medical record data is as follows:

Table 5. Data Transformation

No.	Patient Name	Age	Tension	Diagnosis
1	The Greatest	3	2	21
2	Liang Hong	3	2	22
3	Religious	3	2	1
4	Fatimah Syam	3	2	1
5	Efi Erikawani	3	2	19
6	The Great Wall	3	2	1
7	Minah Samosir	3	3	1
8	Husnah Sitorus	3	2	1
..	..	..	..	..
..	..	..	..	..
100	Zuniar	3	2	22

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Determining the Number of Clusters (K)

The next stage that will be carried out before data processing is to determine the number of clusters. The number of clusters is the number of data groups that will be generated in this study. In this study, the author will group patient medical record data into 3 clusters.

Determining the Initial Centroid

The first iteration process is a stage carried out to obtain patient data groups based on the closest distance of the data to the predetermined cluster center. The first iteration stage consists of several stages, namely determining the initial centroid, calculating the distance, and grouping the data.

At this stage, in determining the initial centroid is taken randomly from the patient data to be processed. The data that will be used as the initial centroid can be seen in table 6. The initial centroid data consists of C1, C2, and C3, here is the initial centroid data table:

Table 6. Determination of Initial Centorid Value

Centroid	Age	Tension	Diagnosis
C1	3	2	1
C2	2	2	22
C3	3	2	15

Calculate the Distance of the Object to the Initial Centroid

To calculate the distance between the data and the centroid, the Euclidean Distance formula is used.*[17], [18], [19], [20]*. This calculation is also performed on patient data 3 to 24. The results of the calculation of the distance of patient data with the three initial cluster centers. The following are the results of iteration 1 in Table 7.

Table 7. Distance of Each Data in Iteration 1

Patient	*Cluster*		
	C1	C2	C3
1	20	1	6
2	21	1	7
3	0	21	14
4	0	21	14
5	18	3	4
6	0	21	14
7	1	21	14
8	0	21	14
..	..	..	..
100	1	20	13

Cluster Grouping Based on Proximity Values

*Clustering*objects by inserting each object into a cluster based on its minimum distance. A data will be a member of a cluster (C1, C2 or C3) that has the smallest distance or value from

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

its cluster center, for example for the first data, the smallest distance is in the 3rd cluster which is 6.164414003, then the first data will be a member of cluster 2. The following are the results of iteration 1 in table 8.

Table 8. Results of Iteration 1 Grouping

Patient	Cluster			Member Cluster
	C1	C2	C3	
1	20	1	6	2
2	21	1	7	2
3	0	21	14	1
4	0	21	14	1
5	18	3	4	2
6	0	21	14	1
7	1	21	14	1
8	0	21	14	1
..	..	..	..	..
24	1	20	13	1

Determining the New Centroid Center

For the second iteration process is a stage carried out to obtain patient data groups based on the closest distance of the data to the new cluster center. Determining the new cluster center is the next stage in the second iteration. The new cluster center is determined based on the grouping of members of each cluster. The average obtained from all cluster groups in the first iteration will be the new centroid center.

Table 9. Results of Determining the New Centroid Center

	New Centroid Center		
cluster 1 = C1	2,857	2,163	1,980
cluster 2 = C2	3,321	2,071	23,536
cluster 3 = C3	2,478	1,826	13,478

Calculating Object Distance to New Centroid

To determine the new centroid center, the next process is to do iteration 2. The two-step iteration is the same as iteration 1. Iteration two is of course calculated using the results of determining the new centroid. The second iteration process will produce a data group in the first data processing. The author uses the Euclidean formula to calculate the distance of the data to the new centroid. This calculation is also carried out on patient data 3 to 24. The results of calculating the distance of patient data with the three new cluster centers. The following are the results of iteration 2 in Table 10.

Table 10. Results of Iteration 2

Patient To	*Cluster*		
	C1	C2	C3
1	19	3	8

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Patient To	*Cluster*		
	C1	C2	C3
2	20	2	9
3	1	23	12
4	1	23	12
5	17	5	6
6	1	23	12
7	1	23	13
8	1	23	12
..	..	..	..
24	1	22	12

Second Iteration Grouping Results

At this stage, the method is the same as the first iteration grouping. Clustering objects by inserting each object into a cluster based on its minimum distance. A data will be a member of a cluster (C1, C2 or C3) that has the smallest distance or value from its cluster center. The following are the results of the second iteration grouping in table 11.

Table 11. Results of Iteration 2 Grouping

Patient to	*Cluster*			Member
	C1	C2	C3	*Cluster*
1	19	3	8	2
2	20	2	9	2
3	1	23	12	1
4	1	23	12	1
5	17	5	6	2
6	1	23	12	1
7	1	23	13	1
8	1	23	12	1
..	..	..	..	..
100	1	22	12	1

K-Means Clustering Results

The K-means iteration process stops when the iteration result does not change from the previous iteration result. The iteration process will continue until the second iteration does not change from the previous iteration value. Based on the cluster results of the patient data, we can conclude that, C1 there are 49 patients, a disease that often occurs with high intensity so that the hospital pays special attention to the disease, namely by providing outdoor counseling, indoor counseling and adding facilities for the condition of the disease. C2 there are 28 patients, a disease that occurs with moderate intensity so that the action that can be taken by the hospital is by means of indoor counseling, C3 there are 23 patients, a disease that occurs with low intensity so that the hospital only gives directions or warnings when the

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

patient is undergoing treatment. In Iteration 2, there is no change in the centroid and clustering results with iteration 1. So it can be said that the iteration process stops until iteration 2 and gets the clustering results, namely, cluster 1 has 49 data, cluster 2 has 28 data and cluster 3 has 23 data. Because in the 1st and 2nd iterations the cluster position does not change, the calculation is stopped and the following graphical results are obtained:

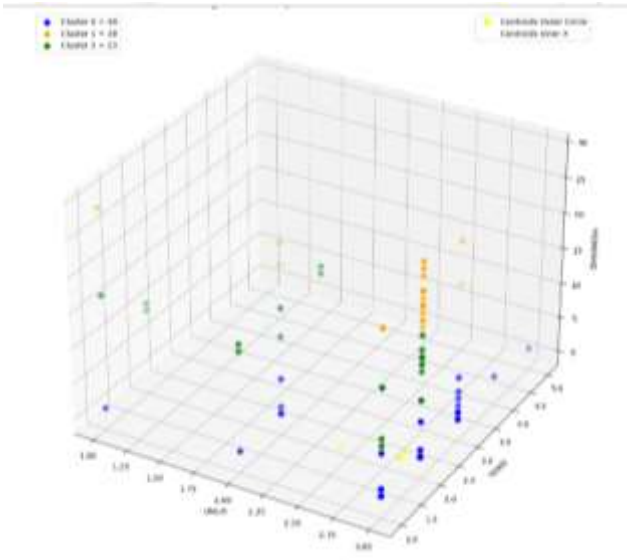

Figure 1. Cluster graph based on the calculations that have been carried out

From 100 data obtained 3 groups, cluster 1 contains 49 patient data cluster 2 contains 28 patient data cluster 3 contains 23 patient data, and the largest group obtained in Cluster 1 contains 49 Patients. It can be seen that in cluster 1, more patients experience the disease with other Clusters.

CONCLUSION

Based on the research results, it was found that grouping diseases using the K-Means Clustering method produced three main clusters, namely Cluster C1 as a disease with high intensity, Cluster C2 with moderate intensity, and Cluster C3 with low intensity. With the clustering method, patients who experience diseases can be grouped into segments with similar characteristics, making it easier for hospitals to understand disease patterns and find more effective healing solutions. Information obtained from this method can also help improve the quality of hospital services, such as drug stock management, facility arrangements, and optimization of operating hours. In addition, the results of this study indicate the need for hospitals to provide more intensive care rooms, especially for elderly patients, so that they get better service. To support the implementation of this method, hospitals are advised to develop a more modern computerized patient data recording system, replacing manual recording. Hospitals also need to improve supporting facilities such as computers and data storage capacity, and provide training to staff to understand the results of clustering analysis. Periodic monitoring and evaluation of clustering results are essential to ensure the accuracy and

relevance of patient segments, including data integration with other systems, such as online-based expert systems, in order to provide more integrated services to patients.

REFERENCES

[1] N. A. Sudibyo *et al.*, "PENERAPAN DATA MINING PADA JUMLAH PENDUDUK," vol. 1, no. 3, pp. 199–207, 2020.

[2] S. N. Pendidikan, U. M. Semarang, and E. Kurniawan, "Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association Rules Dengan Algoritma Clustering," pp. 1–7.

[3] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma," vol. 3, no. 2, pp. 55–66, 2023.

[4] J. Faran and A. Triayudi, "Penerapan Algoritma K-Means Data Mining untuk Clustering Kinerja Karyawan Koperasi," vol. 4, no. 4, pp. 2096–2108, 2024, doi: 10.30865/klik.v4i4.1728.

[5] J. Informasi, J. Wandana, and S. Defit, "Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means," vol. 2, pp. 4–9, 2020, doi: 10.37034/jidt.v2i4.73.

[6] D. Andreswari, R. Efendi, and K. Prastio, "CLUSTERING DATA REKAM MEDIS UNTUK PENENTUAN PENYAKIT DENGAN MENGIMPLEMENTASIKAN METODE FUZZY C-MEANS," vol. 11, no. 1, 2023.

[7] A. Gulo and M. Anis, "Analisa Dan Perancangan Sistem Pengelompokan Data Rekam Medis Pasien Pada Puskesmas Cipayung Depok Menggunakan Metode K-Means Berbasis Web," vol. 2, no. 1, pp. 17–27, 2024.

[8] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26

[9] F. Putri, A. Hasibuan, S. Sumarno, and I. Parlina, "Penerapan K-Means pada Pengelompokan Penjualan Produk Smartphone," vol. 1, no. 1, pp. 15–20, 2021, doi: 10.54259/satesi.v1i1.3.

[10] T. A. Aria, M. Julkarnain, and F. Hamdani, "Penerapan Algoritma K-Means Clustering Untuk Data Obat," vol. 4, no. 1, pp. 649–657, 2023, doi: 10.30865/klik.v4i1.1117.

[11] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.

[12] H. L. Sari and I. Y. Beti, "Penerapan Data Mining Dalam Pengelompokkan Buku Yang Dipinjam Menggunakan Algoritma K-Means," vol. 3, no. 6, pp. 925–933, 2023, doi: 10.30865/klik.v3i6.826.

[13] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 2, no. 1, pp. 30–38, 2021, doi: 10.37148/bios.v2i1.19.

[14] J. Ilmu, S. Informasi, P. Marpaung, I. Pebrian, and W. Putri, "Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means," vol. 6, no. September, pp. 64–70, 2023.

[15] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 1, pp. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.

[16] M. R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *Journal of Software Engineering Ampera*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.

[17] A. Of *et al.*, "Penerapan Algoritma Floyd Warshall dengan Menggunakan Euclidean Distance dalam Menentukan Rute Terbaik," *Jurnal Ilmu Komputer ...*, pp. 312–321, 2023, [Online]. Available: https://jurnal.unity-academy.sch.id/index.php/jirsi/article/view/68%0Ahttps://jurnal.unity-academy.sch.id/index.php/jirsi/article/download/68/51

[18] R. Nuraini, "Implementasi Euclidean Distance dan Segmentasi K-Means Clustering Pada Identifikasi Citra Jenis Ikan Nila," vol. 3, no. 1, pp. 1–8, 2022.

[19] Z. A. Setyawan, "Sistem Pakar Diagnosa Gejala Kecanduan Game Online dengan Metode Certainty Factor & Euclidean Distance Berbasis Web," vol. 1, no. 1, 2023, doi: 10.54259/jdmis.v1i1.1521.

[20] S. Saniah and Edy Victor Haryanto, "Rancang Bangun Aplikasi Pencarian Lokasi Vaksin Pada Puskesmas Di Kota Medan Menggunakan Metode Euclidean Distance Berbasis Android," *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 3, no. 1, pp. 01–11, 2023, doi: 10.55606/juisik.v3i1.365.