



Pengelompokan Penyakit pada Pasien Berdasarkan Usia dengan Metode K-Means Clustering

Maida Andriani^{1*}, Akim Manaor Hara Pardede², Magdalena Simanjuntak³
¹⁻³ STMIK Kaputama Binjai

Alamat: Jl. Veteran No. 4A, Tangsi, Binjai Kota, Kota Binjai, Sumatera Utara 20714

Korespondensi penulis: andrianimaida18@gmail.com

Abstract. *This research aims to cluster disease data based on patient age using the K-Means method at RSUD Dr. RM. Djoelham. In this case study, the clustering method with the K-Means algorithm is used to group patients based on patient age, address and type of disease. With this method, information can be obtained regarding patient grouping patterns based on age at Dr. RM. Djoelham, who helps identify the closest relationships between patient groups and provides insight into the distribution of disease across age groups, regions and types of disease suffered. This research was conducted at RSUD Dr. RM. Djoelham by loading data from patients treated at the hospital. The data used is 1,100 patient data from 2022-2024 which has been recorded by the hospital. This patient data will be analyzed using 3 variables in the research, namely Patient Age (C1), Address (C2), and Type of Disease (C3). With the results, cluster 1 contains 320 data, cluster 2 contains 326 data, and cluster 3 contains 454 data.*

Keywords: *Data Mining; Patient Clustering; K-Means*

Abstrak. Penelitian ini bertujuan untuk melakukan clustering data penyakit berdasarkan usia pasien menggunakan metode K-Means di RSUD Dr. RM. Djoelham. Dalam studi kasus ini, metode clustering dengan algoritma K-Means digunakan untuk mengelompokkan pasien berdasarkan Umur Pasien, Alamat, dan Jenis Penyakit. Dengan metode ini, dapat diperoleh informasi mengenai pola pengelompokan pasien berdasarkan usia di RSUD Dr. RM. Djoelham, yang membantu mengidentifikasi hubungan terdekat antar kelompok pasien dan memberikan wawasan tentang distribusi penyakit di berbagai kelompok usia, wilayah, dan jenis penyakit yang diderita. Penelitian ini dilakukan di RSUD Dr. RM. Djoelham dengan memuat data dari pasien yang dirawat di rumah sakit tersebut. Data yang digunakan adalah 1.100 data pasien dari tahun 2022-2024 yang telah didata oleh pihak rumah sakit. Data pasien ini akan dianalisis menggunakan 3 variabel dalam penelitian, yaitu Umur Pasien (C1), Alamat (C2), dan Jenis Penyakit (C3). Dengan hasil cluster 1 terdapat 320 data, cluster 2 terdapat 326 data, dan cluster 3 terdapat 454 data.

Kata kunci: Data Mining; Clustering Pasien; K-Means

1. PENDAHULUAN

Kesehatan merupakan salah satu aspek fundamental dalam kehidupan manusia, yang memiliki dampak langsung terhadap kualitas hidup individu dan kesejahteraan masyarakat secara keseluruhan. Di era modern ini, peningkatan kualitas pelayanan kesehatan menjadi prioritas utama bagi berbagai negara di dunia, termasuk Indonesia. Salah satu upaya yang dilakukan adalah dengan memanfaatkan teknologi informasi dan analisis data dalam bidang medis untuk mengidentifikasi pola-pola penyakit, yang dapat membantu dalam diagnosis, perawatan, dan pencegahan penyakit. Dalam konteks ini, analisis data kesehatan memiliki peran penting dalam memberikan wawasan yang lebih dalam mengenai distribusi dan karakteristik penyakit di masyarakat.

Penyakit seringkali berkaitan dengan usia individu, namun pola penyakit bisa berbeda antar kelompok usia. Menggunakan metode K-Means Clustering untuk mengelompokkan

pasien berdasarkan usia dapat membantu mengidentifikasi pola penyakit yang berbeda-beda pada kelompok usia tertentu. Hal ini dapat mendukung upaya diagnostik yang lebih baik, perencanaan strategi pengobatan yang lebih spesifik, dan pengembangan model prediktif untuk mengidentifikasi pasien yang rentan terhadap penyakit tertentu berdasarkan usia mereka.

Penerapan metode K-Means Clustering merupakan metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengelompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lain yang telah diperoleh data yang memiliki kesamaan bukan data yang sama tetapi memiliki karakteristik yang sama. (Serasi Ginting & Simanjuntak, 2021)

2. KAJIAN TEORITIS

Data Mining

Data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu database yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Berikut ini adalah beberapa definisi tentang pengertian data mining yang dikemukakan oleh beberapa ahli. Data mining adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga *knowledge discovery*, Menurut (Fitri Larasati dan Andy Sapta, 2017). Data mining memiliki pengertian lain yaitu *knowledge discovery* ataupun *pattern recognition* merupakan suatu istilah yang digunakan untuk mendapatkan pengetahuan yang tersembunyi dari kumpulan data yang berukuran sangat besar. Menambahpengetahuandari data atau informasi yang kita miliki, Menurut Susanto dalam (Setiawan , 2016). Data mining merupakan bidang dari beberapa bidang keilmuan yangmenyatukan teknik dari pembelajaran mesin, pengenalan pola, opici, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yangbesar, Menurut Larose dalam (Rony Setiawan, 2016). Data mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut.

Metode Clustering

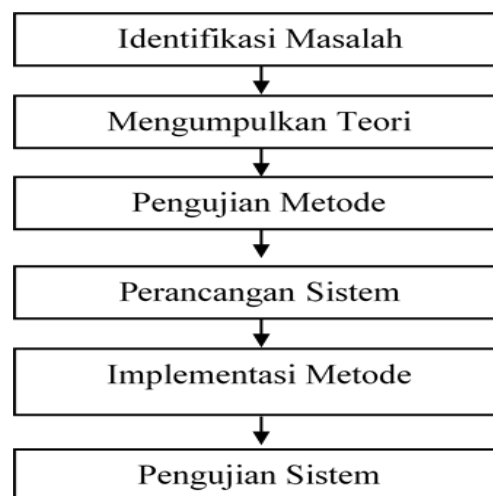
Clustering atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya,

Menurut Widodo dalam (Benri Melpa Metisen dan Herlina Latipa Sari, 2015). Cluster adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam cluster yang sama dan dissimilar terhadap objek-objek yang berbeda cluster. Objek akan dikelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster. Dengan menggunakan clustering ini, kita dapat mengkalsifikasikan daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut data. Dalam data mining, usaha difokuskan pada metode-metode penemuan untuk cluster pada basis data berukuran besar secara efektif, dan efisien dan juga terdapat Beberapa kebutuhan clustering yang dalam data mining meliputi skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda.

3. METODE PENELITIAN

Metodologi penelitian adalah proses atau cara ilmiah untuk mendapatkan data yang akan digunakan untuk keperluan penelitian. Dalam melakukan penelitian pada skripsi ini, penulis mengikuti tahapan metodologi dalam penelitian ini yaitu sebagai berikut :

Tabel 1. metodologi dalam penelitian



Untuk memperjelas struktur metodologi penelitian diatas, maka penulis membuat keterangannya sebagai berikut :

1. Identifikasi Masalah

Tahap ini merupakan tahap awal yang digunakan untuk mengidentifikasi masalah dengan tujuan untuk mengamati dan mencari permasalahan yang sedang dihadapi pada objek penelitian yaitu Pengelompokan penyakit pada pasien berdasarkan usia, dalam tahap ini penulis melakukan riset yang bertujuan untuk mendapatkan sebuah permasalahan yang dihadapi oleh sebuah instansi.

2. Mengumpulkan Teori

Pengumpulan teori-teori yang berhubungan dengan pokok permasalahan seperti teori tentang *Data Mining*, metode yang digunakan dan aplikasi perancangan dari sistem yang diperlukan. Dalam tahap ini, teori dikumpulkan dari beberapa sumber seperti buku-buku, jurnal, artikel dan referesi lainnya.

3. Pengujian Metode

Pada tahap ini peneliti akan menguji metode yang digunakan dalam proses pengelompokan data yang tepat, dengan panduan yang sudah ada pada teori-teori pendukung dari buku-buku maupun jurnal terkait dengan pokok permasalahan.

4. Perancangan Sistem

Pada tahap ini dilakukan perancangan sistem terhadap masalah yang sedang diteliti, bisa berupa tahap untuk merancang alur kerja dari sistem dan juga merancang desain dari tampilan tatap muka (*interface*) dari sistem yang akan dibuat. Rancangan sistem dibuat untuk mempermudah pengguna dalam memahami sistem yang akan dibangun.

5. Implementasi Metode

Mengimplementasikan metode yang sudah diuji sebelumnya dengan rancangan sistem yang telah dibuat serta melakukan pengkodean (*coding*) sesuai dengan bahasa pemrograman yang digunakan untuk membuat sistem tersebut.

4. HASIL DAN PEMBAHASAN

Dalam analisa pengujian metode *K-Means Clustering* dalam sistem pengelompokan data yang digunakan diperlukan data sebagai input proses dan analisa, setelah melakukan riset di Rumah Sakit Umum Djoelham, maka data sebagai pendukung penelitian untuk dianalisa dengan menggunakan metode *K-Means Clustering* adalah sebagai berikut:

Tabel. 2 Data Pendukung Penelitian

No	Nama Pasien	Umur	Alamat	Jenis Penyakit
1	Keysha A	3 Tahun	Binjai Selatan	Observasi Febris
2	Sri S	57 Tahun	Binjai Selatan	Stroke
3	Rosmini	71 Tahun	Binjai Kota	Demam Thypoid
4	Tioria S	70 Tahun	Kec. Berampu	Anemia
5	Muhammad A	23 Tahun	Binjai Selatan	Viral Infection
6	Rivaldi Y	28 tahun	Binjai Kota	Viral Infection
7	Yaya S	59 Tahun	Binjai Kota	Bronchitis
8	Jildun	54 tahun	Bahorok	Viral Infection
9	Sri P	42 Tahun	Binjai Kota	Colic Abdomen
10	Kasiani	58 Tahun	Binjai Timur	Bronchitis
11	Tugimin	56 Tahun	Selesai	CHF ec CAD

No	Nama Pasien	Umur	Alamat	Jenis Penyakit
12	Irfan R	46 Tahun	Binjai Barat	PGK Stage V
13	Listra H	46 Tahun	Binjai Barat	STEMI
14	Maliken S	71 Tahun	Medan	Tuberkulosis Paru
15	Mislan	63 Tahun	Medan	Cystitis Akut
16	Hana S	41 Tahun	Binjai Kota	Septis
17	Tianar T	71 Tahun	Binjai Kota	Hipertensi
18	Mailina	63 Tahun	Deli Serdang	Pneumonia
19	Yusnaningsih	60 Tahun	Binjai Barat	Observasi Febris
20	T. Muhammad S	62 Tahun	Medan	Pneumonia

Dengan mengunakan data penelitian diatas, maka dalam menganalisis metode *Clustering* dengan algoritma *K-Means* dalam mengelompokkan data diatas, berikut langkah-langkahnya:

Menentukan Variabel Yang Digunakan

Berdasarkan data alternatif di atas, variabel yang digunakan adalah sebagai berikut:

- Umur Pasien sebagai Variabel X;
- Alamat sebagai Variabel Y;
- Jenis Penyakit sebagai Variabel Z;

Membuat Nilai Transformasi Data Alternatif

Berdasarkan data alternatif dan variabel yang digunakan maka ada 3 transformasi data yang dapat dibuat, diantaranya yaitu :

- Tabel Transformasi Data Variabel Umur Pasien (X)

Tabel. 3 Transformasi Data Variabel Umur Pasien (X)

No	Umur Pasien	Nilai Transformasi
1	0 – 7	1
2	8 – 15	2
3	16 – 23	3
4	24 – 31	4
5	32 – 39	5
6	40 – 47	6
7	48 – 55	7
8	56 – 63	8
9	64 – 71	9
10	> 72	10

- Tabel Transformasi Data Variabel Alamat (Y)

Tabel. 4 Transformasi Data Variabel Alamat (Y)

No	Alamat	Nilai Transformasi
1	Binjai Selatan	1
2	Binjai Kota	2
3	Kec Berampu	3

4	Bahorok	4
5	Binjai Timur	5
6	Selesai	6
7	Binjai Barat	7
8	Medan	8
9	Deli Sedang	9
10	Langkat	10
11	Binjai Utara	11
12	Stabat	12
13	Rambung Timur	13
14	Mulyo Rejo	14

c. Tabel Transformasi Data Variabel Jenis Penyakit (Z)

Tabel. 5 Transformasi Data Variabel Jenis Penyakit (Z)

No	Jenis Penyakit	Nilai Transformasi
1	Observasi Febris	1
2	Stroke	2
3	Demam Thypoid	3
4	Anemia	4
5	Viral Infection	5
6	Bronchitis	6
7	Colic Abdomen	7
8	CHF ec CAD	8
9	PGK Stage V	9
10	STEMI	10
11	Tuberkulosis Paru	11
12	Cystitis Akut	12
13	Septis	13
14	Hipertensi	14
15	Pneumonia	15
16	UAP	16
17	GE	17
18	Fangiritis Akut	18
19	Vertigo	19

Membuat Data Alternatif Yang Telah Diubah Menjadi Nilai Transformasi Data yang telah diubah menjadi nilai transformasi adalah sebagai berikut :

Tabel. 6 Data Yang Telah Diubah Dengan Nilai Transformasi

No	Alternatif	X	Y	Z
1	A1	1	1	1
2	A2	8	1	2
3	A3	9	2	3
4	A4	9	3	4
5	A5	3	1	5
6	A6	4	2	5
7	A7	8	2	6

No	Alternatif	X	Y	Z
8	A8	7	4	5
9	A9	6	2	7
10	A10	8	5	6
11	A11	8	6	8
12	A12	6	7	9
13	A13	6	7	10
14	A14	9	8	11
15	A15	8	8	12
16	A16	6	2	13
17	A17	9	2	14
18	A18	8	9	15
19	A19	8	7	1
20	A20	8	8	15

1. Menentukan Jumlah *Cluster*

Jumlah *cluster* yang digunakan adalah 3 *cluster* data.

2. Menentukan Pusat *Cluster* (Centroid) Secara Acak (*Random*)

Pusat *cluster* (Centroid) yang digunakan berjumlah 3, yaitu sebagai berikut :

Centroid1 diambil dari data ke- yaitu alternatif A9 : $C1 = (6,2,7)$

Centroid2 diambil dari data ke- yaitu alternatif A5 : $C2 = (3,1,5)$

Centroid3 diambil dari data ke- yaitu alternatif A20 : $C3 = (8,8,15)$

Mengitung Nilai Jarak Antardata Dengan Persamaan *Euclidean Distance* Perhitungan nilai jarak terdekat antardata dengan menggunakan persamaan

$D(ij) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + (X3i - X3j)^2}$. Proses Perhitungan tersebut adalah sebagai berikut :

PROSES ITERASI I

- A1 (1,1,1)

Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$

$$C1 = \sqrt{(1-6)^2 + (1-2)^2 + (1-7)^2} = 7.87$$

$$C2 = \sqrt{(1-3)^2 + (1-1)^2 + (1-5)^2} = 4.47$$

$$C3 = \sqrt{(1-8)^2 + (1-8)^2 + (1-15)^2} = 17.15$$

- A2 (8,1,2)

Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$

$$C1 = \sqrt{(8-6)^2 + (1-2)^2 + (2-7)^2} = 5.48$$

$$C2 = \sqrt{(8-3)^2 + (1-1)^2 + (2-5)^2} = 5.83$$

$$C3 = \sqrt{(8-8)^2 + (1-8)^2 + (2-15)^2} = 14.76$$

- A3 (9,2,3)

Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$

$$C1 = \sqrt{(9-6)^2 + (2-2)^2 + (3-7)^2} = 5$$

$$C2 = \sqrt{(9-3)^2 + (2-1)^2 + (3-5)^2} = 6.40$$

- $$C3 = \sqrt{(9-8)^2 + (2-8)^2 + (3-15)^2} = 13.45$$
- A4 (9,3,4)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(9-6)^2 + (2-2)^2 + (3-7)^2} = 4.36$$
 - $$C2 = \sqrt{(9-3)^2 + (2-1)^2 + (3-5)^2} = 6.40$$
 - $$C3 = \sqrt{(9-8)^2 + (2-8)^2 + (3-15)^2} = 12.12$$
- A5 (3,1,5)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(3-6)^2 + (1-2)^2 + (5-7)^2} = 3.74$$
 - $$C2 = \sqrt{(3-3)^2 + (1-1)^2 + (5-5)^2} = 0$$
 - $$C3 = \sqrt{(3-8)^2 + (1-8)^2 + (5-15)^2} = 13.19$$
- A6 (4,2,5)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(4-6)^2 + (2-2)^2 + (5-7)^2} = 2.83$$
 - $$C2 = \sqrt{(4-3)^2 + (2-1)^2 + (5-5)^2} = 1.41$$
 - $$C3 = \sqrt{(4-8)^2 + (2-8)^2 + (5-15)^2} = 12.33$$
- A7 (8,2,6)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(8-6)^2 + (2-2)^2 + (6-7)^2} = 2.24$$
 - $$C2 = \sqrt{(8-3)^2 + (2-1)^2 + (6-5)^2} = 5.19$$
 - $$C3 = \sqrt{(8-8)^2 + (2-8)^2 + (6-15)^2} = 10.82$$
- A8 (7,4,5)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(7-6)^2 + (4-2)^2 + (5-7)^2} = 3$$
 - $$C2 = \sqrt{(7-3)^2 + (4-1)^2 + (5-5)^2} = 5$$
 - $$C3 = \sqrt{(7-8)^2 + (4-8)^2 + (5-15)^2} = 10.82$$
- A9 (6,2,7)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(6-6)^2 + (2-2)^2 + (7-7)^2} = 0$$
 - $$C2 = \sqrt{(6-3)^2 + (2-1)^2 + (7-5)^2} = 3.74$$
 - $$C3 = \sqrt{(6-8)^2 + (2-8)^2 + (7-15)^2} = 10.20$$
- A10 (8,5,6)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(8-6)^2 + (5-2)^2 + (6-7)^2} = 3.74$$
 - $$C2 = \sqrt{(8-3)^2 + (5-1)^2 + (6-5)^2} = 6.48$$
 - $$C3 = \sqrt{(8-8)^2 + (5-8)^2 + (6-15)^2} = 9.49$$
- A11 (8,6,8)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)
 - $$C1 = \sqrt{(8-6)^2 + (6-2)^2 + (8-7)^2} = 4.58$$
 - $$C2 = \sqrt{(8-3)^2 + (6-1)^2 + (8-5)^2} = 7.68$$
 - $$C3 = \sqrt{(8-8)^2 + (6-8)^2 + (8-15)^2} = 7.28$$
- A12 (6,7,9)
 - Centroid : C1 = (6,2,7); C2 = (3,1,5) dan C3 = (8,8,15)

- $C1 = \sqrt{(6-6)^2 + (7-2)^2 + (9-7)^2} = 5.39$
 $C2 = \sqrt{(6-3)^2 + (7-1)^2 + (9-5)^2} = 7.81$
 $C3 = \sqrt{(6-8)^2 + (7-8)^2 + (9-15)^2} = 6.40$
- A13 (6,7,10)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(6-6)^2 + (7-2)^2 + (10-7)^2} = 5.83$
 $C2 = \sqrt{(6-3)^2 + (7-1)^2 + (10-5)^2} = 8.37$
 $C3 = \sqrt{(6-8)^2 + (7-8)^2 + (10-15)^2} = 5.48$
 - A14 (9,8,11)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(9-6)^2 + (8-2)^2 + (11-7)^2} = 7.81$
 $C2 = \sqrt{(9-3)^2 + (8-1)^2 + (11-5)^2} = 11$
 $C3 = \sqrt{(9-8)^2 + (8-8)^2 + (11-15)^2} = 4.12$
 - A15 (8,8,12)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(8-6)^2 + (8-2)^2 + (12-7)^2} = 8.06$
 $C2 = \sqrt{(8-3)^2 + (8-1)^2 + (12-5)^2} = 11.09$
 $C3 = \sqrt{(8-8)^2 + (8-8)^2 + (12-15)^2} = 3$
 - A16 (6,2,13)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(6-6)^2 + (2-2)^2 + (13-7)^2} = 6$
 $C2 = \sqrt{(6-3)^2 + (2-1)^2 + (13-5)^2} = 8.60$
 $C3 = \sqrt{(6-8)^2 + (2-8)^2 + (13-15)^2} = 6.63$
 - A17 (9,2,14)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(9-6)^2 + (2-2)^2 + (14-7)^2} = 7.62$
 $C2 = \sqrt{(9-3)^2 + (2-1)^2 + (14-5)^2} = 10.86$
 $C3 = \sqrt{(9-8)^2 + (2-8)^2 + (14-15)^2} = 6.16$
 - A18 (8,9,15)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(8-6)^2 + (9-2)^2 + (15-7)^2} = 10.82$
 $C2 = \sqrt{(8-3)^2 + (9-1)^2 + (15-5)^2} = 13.75$
 $C3 = \sqrt{(8-8)^2 + (9-8)^2 + (15-15)^2} = 1$
 - A19 (8,7,1)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(8-6)^2 + (7-2)^2 + (1-7)^2} = 8.06$
 $C2 = \sqrt{(8-3)^2 + (7-1)^2 + (1-5)^2} = 8.77$
 $C3 = \sqrt{(8-8)^2 + (7-8)^2 + (1-15)^2} = 14.03$
 - A20 (8,8,15)
 Centroid : $C1 = (6,2,7)$; $C2 = (3,1,5)$ dan $C3 = (8,8,15)$
 $C1 = \sqrt{(8-6)^2 + (8-2)^2 + (15-7)^2} = 10.20$
 $C2 = \sqrt{(8-3)^2 + (8-1)^2 + (15-5)^2} = 13.19$
 $C3 = \sqrt{(8-8)^2 + (8-8)^2 + (15-15)^2} = 0$

Hasil perhitungan nilai *euclidean distance* pada iterasi I dapat dilihat pada tabel berikut :

Tabel 7 Nilai Euclidean Distance Pada Iterasi I

No	Alternatif	X	Y	Z	C1	C2	C3	Grup
1	A1	1	1	1	7.87	4.47	17.15	2
2	A2	8	1	2	5.48	5.83	14.76	1
3	A3	9	2	3	5	6.40	13.45	1
4	A4	9	3	4	4.36	6.40	12.12	1
5	A5	3	1	5	3.74	0	13.19	2
6	A6	4	2	5	2.83	1.41	12.33	2
7	A7	8	2	6	2.24	5.19	10.82	1
8	A8	7	4	5	3	5	10.82	1
9	A9	6	2	7	0	3.74	10.20	1
10	A10	8	5	6	3.74	6.48	9.49	1
11	A11	8	6	8	4.58	7.68	7.28	1
12	A12	6	7	9	5.39	7.81	6.40	1
13	A13	6	7	10	5.83	8.37	5.48	3
14	A14	9	8	11	7.81	11	4.12	3
15	A15	8	8	12	8.06	11.09	3	3
16	A16	6	2	13	6	8.60	6.63	1
17	A17	9	2	14	7.62	10.86	6.16	3
18	A18	8	9	15	10.82	13.75	1	3
19	A19	8	7	1	8.06	8.77	14.03	1
20	A20	8	8	15	10.20	13.19	0	3

Pada tabel hasil proses iterasi I pengelompokan diatas untuk menentukan grup pengelompokan data dapat ditentukan dengan ketentuan berikut ini :

- Grup 1 didapat dengan membandingkan nilai terkecil dari 3 centroid, jika nilai terkecil terdapat di C1 (Centroid1), maka data masuk pada grup 1.
- Grup 2 didapat dengan membandingkan nilai terkecil dari 3 centroid, jika nilai terkecil terdapat di C2 (Centroid2), maka data masuk pada grup 2.
- Grup 3 didapat dengan membandingkan nilai terkecil dari 3 centroid, jika nilai terkecil terdapat di C3 (Centroid3), maka data masuk pada grup 3.

Selanjutnya melakukan hitungan untuk iterasi II, sebelum melakukan hitungan iterasi II, maka perlu dibuat terlebih dahulu untuk pusat Centroid dari ketiga *Cluster*. Berikut ini adalah 3 Centroid yang baru :

- C1 (11 data) :

Total data X dengan grup 1 = 83, maka $X = \frac{83}{11} = 7.5$

Total data Y dengan grup 1 = 14, maka $Y = \frac{14}{11} = 3.7$

Total data Z dengan grup 1 = 51, maka $Z = \frac{51}{11} = 5.8$

Hasil Centroid1 baru : C1 (7,5;3,7;5,8)

- C2 (3 data) :

Total data X dengan grup 2 = 8, maka $X = \frac{8}{3} = 2,6$

Total data Y dengan grup 2 = 4, maka $Y = \frac{4}{3} = 1,3$

Total data Z dengan grup 2 = 11, maka $Z = \frac{11}{3} = 3,6$

Hasil Centroid2 : C2 (2,6;1,3;3,6)

- C3 (6 data):

Total data X dengan grup 3 = 48, maka $X = \frac{48}{6} = 8$

Total data Y dengan grup 3 = 42, maka $Y = \frac{42}{6} = 7$

Total data Z dengan grup 3 = 77, maka $Z = \frac{77}{6} = 12,8$

Hasil Centroid3 : C3 (8;7;12,8)

PROSES ITERASI II

- A1 (1,1,1)

Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(1 - 7,5)^2 + (1 - 3,7)^2 + (1 - 5,8)^2} = 8,53$$

$$C2 = \sqrt{(1 - 2,6)^2 + (1 - 1,3)^2 + (1 - 3,6)^2} = 3,07$$

$$C3 = \sqrt{(1 - 8)^2 + (1 - 7)^2 + (1 - 12,8)^2} = 15,00$$

- A2 (8,1,2)

Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(8 - 7,5)^2 + (1 - 3,7)^2 + (2 - 5,8)^2} = 4,79$$

$$C2 = \sqrt{(8 - 2,6)^2 + (1 - 1,3)^2 + (2 - 3,6)^2} = 5,64$$

$$C3 = \sqrt{(8 - 8)^2 + (1 - 7)^2 + (2 - 12,8)^2} = 12,36$$

- A3 (9,2,3)

Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(9 - 7,5)^2 + (2 - 3,7)^2 + (3 - 5,8)^2} = 3,6$$

$$C2 = \sqrt{(9 - 2,6)^2 + (2 - 1,3)^2 + (3 - 3,6)^2} = 6,47$$

$$C3 = \sqrt{(9 - 8)^2 + (2 - 7)^2 + (3 - 12,8)^2} = 11,05$$

- A4 (9,3,4)

Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(9 - 7,5)^2 + (3 - 3,7)^2 + (4 - 5,8)^2} = 2,45$$

$$C2 = \sqrt{(9 - 2,6)^2 + (3 - 1,3)^2 + (4 - 3,6)^2} = 6,62$$

$$C3 = \sqrt{(9 - 8)^2 + (3 - 7)^2 + (4 - 12,8)^2} = 9,72$$

- A5 (3,1,5)

Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(3 - 7,5)^2 + (1 - 3,7)^2 + (5 - 5,8)^2} = 5,31$$

$$C2 = \sqrt{(3 - 2,6)^2 + (1 - 1,3)^2 + (5 - 3,6)^2} = 1,49$$

$$C3 = \sqrt{(3 - 8)^2 + (1 - 7)^2 + (5 - 12,8)^2} = 11,04$$

- A6 (4,2,5)

Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(4 - 7,5)^2 + (2 - 3,7)^2 + (5 - 5,8)^2} = 3,97$$

- $$C2 = \sqrt{(4 - 2,6)^2 + (2 - 1,3)^2 + (5 - 3,6)^2} = 2,1$$

$$C3 = \sqrt{(4 - 8)^2 + (2 - 7)^2 + (5 - 12,8)^2} = 10,09$$
- A7 (8,2,6)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(8 - 7,5)^2 + (2 - 3,7)^2 + (6 - 5,8)^2} = 1,78$$

$$C2 = \sqrt{(8 - 2,6)^2 + (2 - 1,3)^2 + (6 - 3,6)^2} = 5,95$$

$$C3 = \sqrt{(8 - 8)^2 + (2 - 7)^2 + (6 - 12,8)^2} = 8,44$$
- A8 (7,4,5)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(7 - 7,5)^2 + (4 - 3,7)^2 + (5 - 5,8)^2} = 0,99$$

$$C2 = \sqrt{(7 - 2,6)^2 + (4 - 1,3)^2 + (5 - 3,6)^2} = 5,35$$

$$C3 = \sqrt{(7 - 8)^2 + (4 - 7)^2 + (5 - 12,8)^2} = 8,41$$
- A9 (6,2,7)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(6 - 7,5)^2 + (2 - 3,7)^2 + (7 - 5,8)^2} = 2,56$$

$$C2 = \sqrt{(6 - 2,6)^2 + (2 - 1,3)^2 + (7 - 3,6)^2} = 4,86$$

$$C3 = \sqrt{(6 - 8)^2 + (2 - 7)^2 + (7 - 12,8)^2} = 7,91$$
- A10 (8,5,6)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(8 - 7,5)^2 + (5 - 3,7)^2 + (6 - 5,8)^2} = 1,41$$

$$C2 = \sqrt{(8 - 2,6)^2 + (5 - 1,3)^2 + (6 - 3,6)^2} = 6,97$$

$$C3 = \sqrt{(8 - 8)^2 + (5 - 7)^2 + (6 - 12,8)^2} = 7,1$$
- A11 (8,6,8)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(8 - 7,5)^2 + (6 - 3,7)^2 + (8 - 5,8)^2} = 3,22$$

$$C2 = \sqrt{(8 - 2,6)^2 + (6 - 1,3)^2 + (8 - 3,6)^2} = 8,4$$

$$C3 = \sqrt{(8 - 8)^2 + (6 - 7)^2 + (8 - 12,8)^2} = 4,9$$
- A12 (6,7,9)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(6 - 7,5)^2 + (7 - 3,7)^2 + (9 - 5,8)^2} = 4,84$$

$$C2 = \sqrt{(6 - 2,6)^2 + (7 - 1,3)^2 + (9 - 3,6)^2} = 8,56$$

$$C3 = \sqrt{(6 - 8)^2 + (7 - 7)^2 + (9 - 12,8)^2} = 4,29$$
- A13 (6,7,10)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(6 - 7,5)^2 + (7 - 3,7)^2 + (10 - 5,8)^2} = 5,55$$

$$C2 = \sqrt{(6 - 2,6)^2 + (7 - 1,3)^2 + (10 - 3,6)^2} = 9,22$$

$$C3 = \sqrt{(6 - 8)^2 + (7 - 7)^2 + (10 - 12,8)^2} = 3,44$$
- A14 (9,8,11)
Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)

$$C1 = \sqrt{(9 - 7,5)^2 + (8 - 3,7)^2 + (11 - 5,8)^2} = 6,91$$

$$C2 = \sqrt{(9 - 2,6)^2 + (8 - 1,3)^2 + (11 - 3,6)^2} = 11,86$$

$$C3 = \sqrt{(9 - 8)^2 + (8 - 7)^2 + (11 - 12,8)^2} = 2,29$$
- A15 (8,8,12)

- Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)
- $$C1 = \sqrt{(8 - 7,5)^2 + (8 - 3,7)^2 + (12 - 5,8)^2} = 7,56$$
- $$C2 = \sqrt{(8 - 2,6)^2 + (8 - 1,3)^2 + (12 - 3,6)^2} = 12,02$$
- $$C3 = \sqrt{(8 - 8)^2 + (8 - 7)^2 + (12 - 12,8)^2} = 1,28$$
- A16 (6,2,13)
- Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)
- $$C1 = \sqrt{(6 - 7,5)^2 + (2 - 3,7)^2 + (13 - 5,8)^2} = 7,55$$
- $$C2 = \sqrt{(6 - 2,6)^2 + (2 - 1,3)^2 + (13 - 3,6)^2} = 10,2$$
- $$C3 = \sqrt{(6 - 8)^2 + (2 - 7)^2 + (13 - 12,8)^2} = 5,39$$
- A17 (9,2,14)
- Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)
- $$C1 = \sqrt{(9 - 7,5)^2 + (2 - 3,7)^2 + (14 - 5,8)^2} = 8,51$$
- $$C2 = \sqrt{(9 - 2,6)^2 + (2 - 1,3)^2 + (14 - 3,6)^2} = 12,23$$
- $$C3 = \sqrt{(9 - 8)^2 + (2 - 7)^2 + (14 - 12,8)^2} = 5,24$$
- A18 (8,9,15)
- Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)
- $$C1 = \sqrt{(8 - 7,5)^2 + (9 - 3,7)^2 + (15 - 5,8)^2} = 10,63$$
- $$C2 = \sqrt{(8 - 2,6)^2 + (9 - 1,3)^2 + (15 - 3,6)^2} = 14,78$$
- $$C3 = \sqrt{(8 - 8)^2 + (9 - 7)^2 + (15 - 12,8)^2} = 2,97$$
- A19 (8,7,1)
- Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)
- $$C1 = \sqrt{(8 - 7,5)^2 + (7 - 3,7)^2 + (1 - 5,8)^2} = 5,85$$
- $$C2 = \sqrt{(8 - 2,6)^2 + (7 - 1,3)^2 + (1 - 3,6)^2} = 8,27$$
- $$C3 = \sqrt{(8 - 8)^2 + (7 - 7)^2 + (1 - 12,8)^2} = 11,8$$
- A20 (8,8,15)
- Centroid : C1 (7,5;3,7;5,8); C2 (2,6;1,3;3,6); C3 (8;7;12,8)
- $$C1 = \sqrt{(8 - 7,5)^2 + (8 - 3,7)^2 + (15 - 5,8)^2} = 10,16$$
- $$C2 = \sqrt{(8 - 2,6)^2 + (8 - 1,3)^2 + (15 - 3,6)^2} = 14,28$$
- $$C3 = \sqrt{(8 - 8)^2 + (8 - 7)^2 + (15 - 12,8)^2} = 2,42$$

Dari hasil perhitungan nilai *euclidean distance* pada iterasi I di atas dapat dilihat pada tabel berikut:

Tabel 8 Nilai *Euclidean Distance* Pada Iterasi II

No	Alternatif	X	Y	Z	C1	C2	C3	Grup
1	A1	1	1	1	8,53	3,07	15	2
2	A2	8	1	2	4,79	5,64	12,36	1
3	A3	9	2	3	3,36	6,47	11,05	1
4	A4	9	3	4	2,45	6,62	9,72	1
5	A5	3	1	5	5,31	1,49	11,04	2
6	A6	4	2	5	3,97	2,1	10,09	2
7	A7	8	2	6	1,78	5,95	8,44	1
8	A8	7	4	5	0,99	5,35	8,41	1
9	A9	6	2	7	2,56	4,86	7,91	1
10	A10	8	5	6	1,41	6,97	7,1	1

No	Alternatif	X	Y	Z	C1	C2	C3	Grup
11	A11	8	6	8	3,22	8,4	4,9	1
12	A12	6	7	9	4,84	8,56	4,29	3
13	A13	6	7	10	5,55	9,22	3,44	3
14	A14	9	8	11	6,91	11,86	2,29	3
15	A15	8	8	12	7,56	12,02	1,28	3
16	A16	6	2	13	7,55	10,2	5,39	3
17	A17	9	2	14	8,51	12,23	5,24	3
18	A18	8	9	15	10,63	14,78	2,97	3
19	A19	8	7	1	5,85	8,27	11,8	1
20	A20	8	8	15	10,16	14,28	2,42	3

Selanjutnya mengecek kesamaan antara grup awal dengan grup baru, yaitu pada grup data iterasi I dengan data pada iterasi II, proses ini untuk memeriksa apakah terjadi perubahan pada grup *cluster* atau tidak, berikut perbandingan kesamaan antargrup awal dengan grup baru:

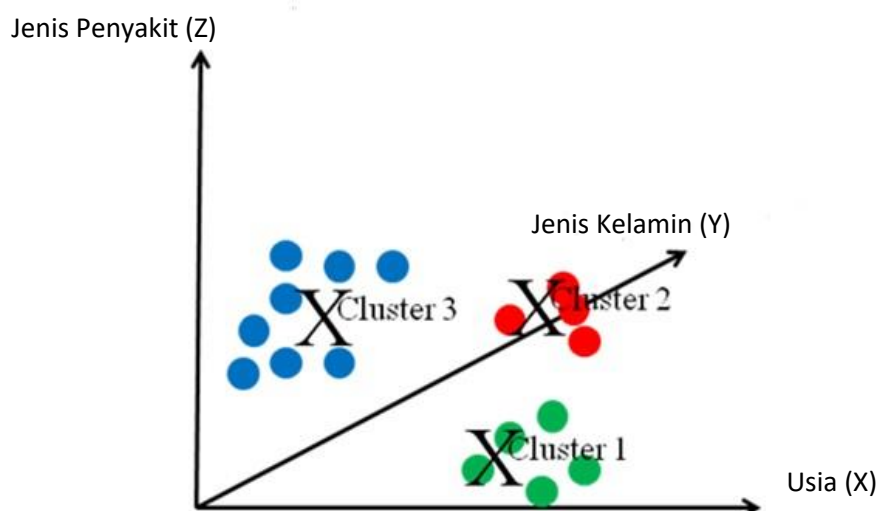
Tabel 9 Perbandingan Grup Iterasi I dan Iterasi II

No	Alternatif	Grup Pada Iterasi I	Grup Pada Iterasi II
1	A1	2	2
2	A2	1	1
3	A3	1	1
4	A4	1	1
5	A5	2	2
6	A6	2	2
7	A7	1	1
8	A8	1	1
9	A9	1	1
10	A10	1	1
11	A11	1	1
12	A12	1	3
13	A13	3	3
14	A14	3	3
15	A15	3	3
16	A16	1	3
17	A17	3	3
18	A18	3	3
19	A19	1	1
20	A20	3	3

Hasil perbandingan dari grup awal (iterasi I) dan grup baru (iterasi II) sama, perhitungan iterasi selesai.

Membuat Grafik *Clustering*

Grafik *Clustering* merupakan grafik yang menampilkan titik-titik koordinat dari sebuah data setelah proses pengelompokan data. Dengan menggunakan aplikasi pemrograman MATLAB, berikut ini adalah hasil grafik *Clustering* pengelompokan data:



Gambar 1 Grafik *Clustering* Data Dengan Metode *K-Means Clustering*

Dari 20 data yang diolah menjadi alternatif pengelompokan data menggunakan metode *Clustering* pada penelitian ini, diperoleh 3 grup; grup 1 (*Cluster 1*) terdapat 9 data, grup 2 (*Cluster 2*) terdapat 3 data dan grup 3 (*Cluster 3*) terdapat 8 data.

1. *Cluster 1* (7,5;3,7;5,8):

Data dapat dikelompokkan berdasarkan Usia (X) “56 – 63” dengan Alamat (Y) di “Bahorok” dan Jenis Penyakit (Z) “Bronchitis”.

2. *Cluster 2* (2,6;1,3;3,6):

Data dapat dikelompokkan berdasarkan Usia (X) “16 – 23” dengan Alamat (Y) di “Binjai Kota” dan Jenis Penyakit (Z) “Anemia”.

3. *Cluster 3* (8;7;12,8):

Data dapat dikelompokkan berdasarkan Usia (X) “56 – 63” dengan Alamat (Y) di “Binjai Barat” dan Jenis Penyakit (Z) “Septis”.

5. KESIMPULAN

Dari hasil pengujian dan analisis yang dilakukan, maka diambil kesimpulan sebagai berikut:

1. Pada kelompok usaha grup 1 dengan jumlah data 320, dapat dikelompokkan berdasarkan Umur Pasien “48 – 55” dengan Alamat “Langkat” dan Jenis Penyakit “Viral Infection”.
2. Pada kelompok usaha grup 2 dengan jumlah data 326, dapat dikelompokkan berdasarkan Umur Pasien “48 – 55” dengan Nama Alamat “Kec Berampu” dan Jenis Penyakit “Viral Infection”.
3. Pada kelompok usaha grup 3 dengan jumlah data 454, dapat dikelompokkan berdasarkan Umur Pasien “48 – 55” dengan Alamat “Binjai Timur” dan Jenis Penyakit “Septis”.

DAFTAR REFERENSI

- Amelia, R., Martanto, M., & Bahtiar, A. (2022). Penerapan algoritma K-Means clustering pada pengelompokan pasangan usia subur peserta KB di Kabupaten Cirebon. *Jurnal Teknologi Technoscientia*, 15(1), 9–14. <https://doi.org/10.34151/technoscientia.v15i1.3849>
- Bayu Prasetyo, R., Agus Pranoto, Y., & Primaswara Prasetya, R. (2023). Implementasi data mining menggunakan algoritma K-Means clustering penyakit pasien rawat jalan pada Klinik Dr. Atirah Desa Sioyong, Sulteng. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2144–2151. <https://doi.org/10.36040/jati.v7i4.7419>
- Fauzi, J. R. (2020). Algoritma dan flowchart dalam menyelesaikan suatu masalah disusun oleh Universitas Janabadra Yogyakarta 2020. *Jurnal Teknik Informatika*, 20330044, 4–6.
- Fauziah, R., & Purnamasari, A. I. (2023). Implementasi algoritma K-Means pada kasus kekerasan anak dan perempuan berdasarkan usia. *Hello World Jurnal Ilmu Komputer*, 2(1), 34–41. <https://doi.org/10.56211/helloworld.v2i1.232>
- Fayyad, U. (1996). *Advances in knowledge discovery and data mining*. California: AAAI Press.
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma C4.5 (Studi kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2), 130–138.
- Iswavigra, D. U., Defit, S., & Nurcahyo, G. W. (2021). Data mining dalam pengelompokan penyakit pasien dengan metode K-Medoids. *Jurnal Informasi Dan Teknologi*, 3, 181–189. <https://doi.org/10.37034/jidt.v3i4.150>
- Khasanah, D. N. (2022). The risk factors of hypertension in Indonesia (Data study of Indonesian family life survey 5). *Journal of Public Health Research and Community Health Development*, 5(2), 80. <https://doi.org/10.20473/jphrecode.v5i2.27923>
- Parasian D. P, S., & Irene Sri, M. (2017). Klusterisasi pola penyebaran penyakit pasien berdasarkan usia pasien dengan menggunakan K-Means clustering. *Jurnal TIMES*, 6(2), 22–25. <http://ejournal.stmik-time.ac.id/index.php/jurnalTIMES/article/view/584>
- Rahmah, L. (2017). Perbandingan hasil penggerombolan K-Means, Fuzzy K-Means, dan Two Step Clustering. *Jurnal Pendidikan Matematika*, 2(1), 39. <https://doi.org/10.18592/jpm.v2i1.1166>
- Rosaly, R., & Prasetyo, A. (2020). Flowchart beserta fungsi dan simbol-simbol. *Journal of Chemical Information and Modeling*, 2(3), 5–7.
- Serasi Ginting, B., & Simanjuntak, M. (2021). Pengelompokan penyakit pada pasien berdasarkan usia dengan metode K-Means clustering (Studi kasus: Puskesmas Bahorok). *ALGORITMA: Jurnal Ilmu Komputer Dan Informatika*, 6341(November), 2.
- Setiawan, R. (2021). Flowchart adalah: Fungsi, jenis, simbol, dan contohnya. Dicoding.
- Sibuea, M. L., & Safta, A. (2017). Pemetaan siswa berprestasi menggunakan metode K-Means clustering. *Jurteksi*, 4(1), 85–92. <https://doi.org/10.33330/jurteksi.v4i1.28>

Sinaga, A. V., Saifullah, & Hardinata, J. T. (2021). Penerapan algoritma backpropagation dalam memprediksi kebutuhan blangko sertifikat tanah pada Kantor BPN Kota Pematangsiantar. *TIN: Terapan Informatika Nusantara*, 1(4). <https://doi.org/doi:https://ejurnal.seminar-id.com/index.php/tin>

Swastati, L. W. (2017). Pengenalan penyakit pada manusia berbasis Android menggunakan metode sequential search. *Jurnal Mahasiswa Teknik Informatika (JATI)*, 1(1), 277–282.