

# Penerapan Algoritma K-Means Clustering Untuk Klasterisasi Penyakit Pasien

Alpazri Lingga

Program Studi Teknik Informatika, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Budi Darma, Medan, Indonesia

Email: alpazrilinggakampus@gmail.com

**Abstrak**—Rs. Kota Subulussalam, merupakan pusat rumah sakit umum daerah kota Subulussalam Aceh yang memang sudah dikenal dan dipercaya oleh masyarakat setempat. Berbagai macam jenis penyakit pasien yang ada di rumah sakit umum ini, maka dari itu, tujuan dari penelitian ini untuk mengklasterisasi pasien pada Rs. Kota Subulussalam, dengan permasalahan yang sedang terjadi pada rumah sakit tersebut yaitu pencarian data pasien dan juga dalam memprediksi tingkat jumlah pasien pada rumah sakit kota Subulussalam. Sehingga pegawai dan pihak manajemen pada rumah sakit tersebut, berkesulitan dalam mengelompokkan data pasien. Berdasarkan permasalahan diatas, Maka dilakakukan sistem Klasterisasi pasien dengan menggunakan Algoritma K-Means. Proses clasterisasi dibagi menjadi tiga kategori yaitu self care, intermediate care dan total care. Dengan klasterisasi pasien dapat menentukan banyak dan jenistenaga yang dibutuhkan dan menentukan nilai produktivitas sehingga jumlah perawat dengan kebutuhan pasien dapat seimbang. Dengan menggunakan algoritma K-Means Clustering ini penulis akan mengklasterisasikan data penyakit pasien rumah sakit umum daerah kota subulussalam, dengan harapan agar kiranya algoritma ini dapat menjadi solusi permasalahan selama ini, sehingga akan memudahkan pihak-pihak yang terkait dalam pengembangan klasterisasian yang penulis susun ini.

**Kata Kunci:** Data Mining; K-Means; Klasterisasi

**Abstract**—Abstract-Rs. Subulussalam City, is the centre of the regional general hospital of Subulussalam city Aceh which is already known and trusted by the local community. Various types of patient diseases that exist in this public hospital, therefore, the purpose of this study is to cluster patients at Rs. Kota Subulussalam, with the problems that are happening in the hospital, namely data retrieval. Subulussalam City, with the problems that are happening in the hospital, namely searching for patient data and also in predicting the level of the number of patients in the Subulussalam city hospital. So that employees and management at the hospital, have difficulty in classifying patient data. Based on the above problems, a patient clustering system was carried out using the K-Means Algorithm. The clustering process is divided into three categories namely self care, intermediate care and total care. With the clustering of patients can determine the number and type of staff needed and determine the value of productivity so that the number of nurses with patient needs can be balanced. By using the K-Means Clustering algorithm, the author will cluster the patient disease data of the subulussalam city public hospital, with the hope that this algorithm can be a solution to the problem so far, so that it will facilitate the parties involved in the development of the clustering that the author compiled.

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

## 1. PENDAHULUAN

Clustering adalah metode yang digunakan dalam data mining. Data mining adalah sistem yang mengambil dan mengelompokkan informasi yang memiliki jenis informasi yang sama antara data yang diterima dengan data lainnya. Keistimewaan dari teknologi information mining ini adalah tidak memiliki arah (solo). Artinya, teknik ini berjalan tanpa pemrosesan informasi atau instruktur dan tidak menggunakan target output[1][2][3].

Rs. Kota Subulussalam, merupakan pusat rumah sakit umum daerah kota Subulussalam Aceh yang memang sudah dikenal dan dipercaya oleh masyarakat setempat. Berbagai macam jenis penyakit pasien yang ada di rumah sakit umum ini, maka dari itu, tujuan dari penelitian ini untuk mengklasterisasi penyakit pasien pada Rs. Kota Subulussalam. Dengan permasalahan yang sedang terjadi pada rumah sakit tersebut yaitu pencarian data penyakit pasien yang dimana data penyakit pasien di rumah sakit umum kota subulussalam masih bercampur dengan seluruh jenis diagnosa penyakit, sehingga menyulitkan tamu atau kunjungan keluarga pasien dalam mencari keberadaan pasien.

Berdasarkan permasalahan diatas, Maka dilakakukan sistem Klasterisasi pasien dengan menggunakan Algoritma K-Means. Algoritma K-Means merupakan salah satu dari algoritma data mining yang well known di kalangan peneliti. Jika dibandingkan dengan algoritma klasterisasi lainnya, k-Means juga masih cukup tangguh di berbagai jenis data[4][5]. Proses clasterisasi dibagi menjadi tiga kategori yaitu self care, intermediate care dan total care. Dengan adanya sistem ini dapat memberikan manfaat kemudahan untuk menganalisis dan mengklasterisasi pasien pada Rs. Kota Subulussalam, menentukan dan mengklasterisasikan jenis-jenis penyakit pasien. Algoritma analisis cluster KMeans pada dasarnya dapat diterapkan pada masalah pemahaman jenis penyakit pasien dan identifikasi pasien. Anda juga dapat menggunakan algoritma KMeans ini untuk mengurangi jumlah pasien dari jumlah yang banyak menjadi jenis atau karakteristik penyakit pasien. Mari kita lihat masing-masing kelompok. Metode pengumpulan informasi adalah observasi dan wawancara.

Berdasarkan permasalahan di atas, penelitian ini menggunakan metode K-Means untuk mendapatkan informasi pengelompokan penyakit “akut” dan “non-akut” yang banyak diderita oleh pasien di Puskesmas Kajen Pekalongan. Hasilnya dapat dijadikan sebagai sumber atau dasar pendidikan kesehatan oleh dinas kesehatan

setempat. K-Means merupakan salah satu algoritma data mining yang dapat digunakan untuk mengelompokkan/cluster data. Ada banyak pendekatan untuk membuat cluster, seperti membuat aturan yang mendikte keanggotaan dalam grup yang sama berdasarkan kesamaan antar anggota[6][7].

Adapun alasan penulis mengajukan atau membuat judul seperti ini. Di karenakan laporan klasterisasian / Cluster-Cluster di rumah sakit umum daerah kota subulussalam yang belum optimal atau data pasien dan diagnosa penyakit yang masih bercampur dengan bermacam-macam jenis diagnosa lainnya, sehingga tidak efektif jika tidak diperbaiki atau tidak melakukan perubahan sistem. Dengan adanya penelitian pengklasterisasian pasien agar nanti pihak dari rumah sakit dapat mengidentifikasi data penyakit pasien. Adapun referensi dalam penelitian ini adalah penelitian yang sudah pernah dilakukan oleh peneliti-peneliti sebelumnya, dan diantaranya penelitian yang ditulis oleh Kiki Fatmawati, Agus Perdana Windarto, di publis pada Juli 2018 dengan judul *Data mining: Penerapan RapidMiner Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi*[8].

## 2. METODOLOGI PENELITIAN

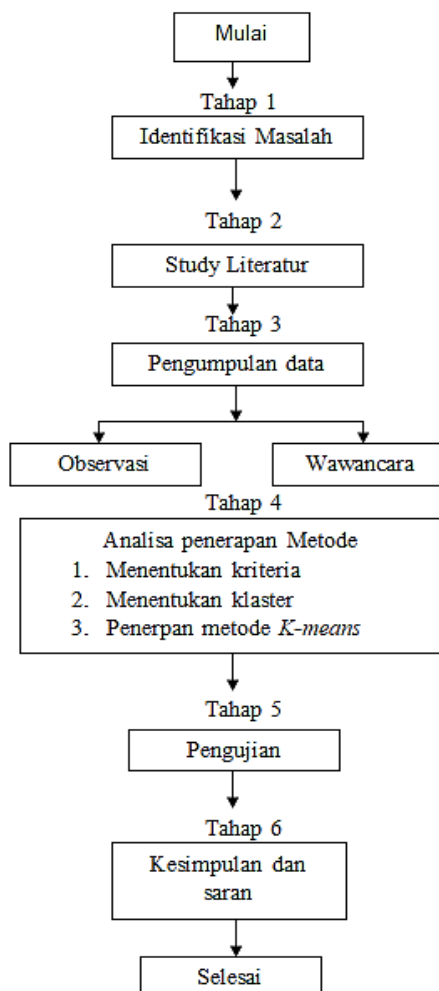
### 2.1 Tahapan Penelitian

Metode survei adalah serangkaian kegiatan, prosedur, atau tahapan yang dilakukan saat melakukan survei. Metode survei dirancang untuk membuat survei ini lebih konseptual dan tepat sasaran untuk tujuan survei. Dalam tahap ini akan dijelaskan mengenai tahapan-tahapan penelitian yang harus dilakukan agar mempermudah dalam penyelesaian terhadap permasalahan yang ada, serta melakukan pengumpulan data yang berkaitan dengan penelitian untuk memperoleh data-data dan informasi yang dibutuhkan. Dalam mengumpulkan data dan informasi yang dibutuhkan penulis melakukan penelitian secara langsung pada Rumah Sakit Umum Daerah (RSUD) Kota Subulussalam Aceh dengan melakukan *observasi*, riset dan wawancara dengan pihak perusahaan yang berkaitan dengan klasterisasi jenis-jenis penyakit pasien.

Dalam kegiatan *observasi* dan riset penulis melakukan wawancara dengan mempertanyakan permasalahan yang terjadi dalam penelitian, yaitu tentang bagaimana proses klasterisasi jenis-jenis penyakit pasien, serta apa saja kriteria yang harus dipenuhi dalam menentukan klasterisasi jenis-jenis penyakit pasien, yang nantinya akan menjadi ketentuan dalam proses penilaian. Setelah melakukan wawancara penulis mendapatkan data dan informasi berupa data pasien yang akan digunakan dalam penelitian serta kriteria penilaian dan bobot nilai yang akan di berikan pada masing-masing kriteria. Setelah mendapatkan informasi dan data yang dibutuhkan, data-data yang diperoleh nantinya akan digunakan dalam menyelesaikan penelitian ini. Dibantu dengan menggunakan dengan menerapkan metode *K-Means*. Dibawah ini merupakan alur dari tahapan penelitian serta sistem pengabungan informasi yang dipakai, yaitu sebagai berikut:

1. Identifikasi masalah  
Yaitu uraian masalah yang melatar belakangi pembuatan skripsi ini tentang permasalahan dalam menentukan penng Klasterisasi penyakit pasien pada Rumah Sakit Kota Subulussalam Aceh.
2. *Study Literatur*  
Yaitu proses pengumpulan bahan-bahan referensi untuk memperoleh informasi yang dibutuhkan dalam pembuatan skripsi ini, baik dari buku, artikel, makalah, jurnal dan situs internet.
3. Pengumpulan Data  
Yaitu pengumpulan data dengan meneliti langsung pada Rumah Sakit Kota Subulussalam Aceh. Dengan cara melibatkan pihak-pihak yang terkait. Hal ini dilakukan untuk pengumpulan data dan informasi yang berhubungan dengan permasalahan. Yaitu dengan cara *observasi* dan wawancara:
  - a. *Observasi* (Pengamatan Langsung)  
Yaitu penulis melakukan pengamatan secara langsung ke perusahaan atau tempat yang diteliti untuk mendapatkan informasi-informasi yang lebih banyak tentang perusahaan, yang nantinya akan digunakan dalam penelitian ini.
  - b. *Interview* (Wawancara)  
Wawancara merupakan teknik pengumpulan data yang dilakukan dengan cara berkomunikasi atau mewawancarai secara langsung kepada pihak perusahaan yang menangani penentuan Klasterisasi pada Rumah Sakit Kota Subulussalam Aceh..
4. Analisa Penerapan Metode  
Pada tahap ini penulis melakukan cluster data pada Pasien Rumah Sakit Kota Subulussalam Aceh. dengan metode *K-Means*. Dalam *clustering* Klasterisasi pasien menggunakan algoritma *K-Means*.
5. Hasil Pengujian  
Peneliti menyimpulkan suatu kesimpulan berdasarkan hasil pengujian melalui perhitungan dan perancangan aplikasi yang dilakukan. Apakah hasil yang diperoleh sesuai dengan maksud dan tujuan penelitian dalam menyelesaikan suatu permasalahan yang dibuat.
6. Kesimpulan dan Saran  
Yaitu merupakan tahapan akhir dari penulisan skripsi ini, dengan memberi kesimpulan berdasarkan hasil yang diperoleh dari penelitian ini. Apakah penelitian ini mampu memecahkan permasalahan untuk mendukung suatu

keputusan dalam menentukan Klasterisasi pasien. Sedangkan saran yang dibuat dapat digunakan sebagai masukan bagi perusahaan untuk dapat dikembangkan lebih lanjut



**Gambar 1.** Tahapan Penelitian

## 2.2 Data mining

Penambangan data adalah konsep yang digunakan untuk menemukan nilai tersembunyi dalam kumpulan data. Penambangan informasi adalah proses semi-otomatis yang menggunakan statistik, matematika, kecerdasan buatan, dan teknologi AI untuk mengekstrak dan mengidentifikasi informasi pengetahuan yang berpotensi berguna yang disimpan dalam kumpulan data besar. Data mining adalah proses penggalian atau pengambilan informasi baru dengan mencari pola atau aturan tertentu dalam sejumlah besar informasi yang diharapkan dapat mengatasi kriteria tersebut. Ada beberapa metode penambangan informasi itu sendiri, salah satunya adalah klasifikasi. Teknik kompleks ini nantinya akan mengidentifikasi dan mengekstrak data yang berguna dari kumpulan data besar. Contoh sederhana adalah membaca data di buku telepon, misalnya. Setelah dibaca, bisa disebut eksposisi, karena memberikan informasi bahwa mayoritas orang bernama Agus tinggal di Tangerang[9][10][11].

Oleh karena itu, dari penjelasan di atas dapat disimpulkan bahwa data mining adalah proses penggalian data dari data dalam jumlah yang sangat besar dengan menggunakan metode statistik dan matematis dengan menggunakan teknologi kecerdasan buatan terbaru. Data mining, juga dikenal sebagai pengenalan pola, adalah metode pemrosesan data untuk menemukan pola tersembunyi dalam data yang diproses. Data yang diolah dengan teknik data mining ini menghasilkan pengetahuan baru yang diperoleh dari data lama. Hasil pengolahan data dapat digunakan untuk pengambilan keputusan di masa yang akan datang.

Menurut para ahli, tujuan dari data mining ini adalah untuk mengekstrak dan mengidentifikasi data untuk informasi spesifik yang terkait dengan database besar atau data besar. Data mining merupakan bagian dari proses KDD (penemuan pengetahuan dalam database) dan terdiri dari beberapa fase seperti pemilihan data, preprocessing, transfer, data mining, dan evaluasi hasil. Teknologi data mining secara garis besar dapat dibagi menjadi dua kelompok: validasi dan penemuan. Metode verifikasi meliputi metode statistik seperti goodness of fit dan analisis varians. Sementara itu, metode discovery dibagi lagi menjadi model prediktif dan model deskriptif. Model prediksi menggunakan hasil yang diketahui dari berbagai data untuk memprediksi data. Model prediktif dapat dibangun di

atas penggunaan data historis lainnya. Model deskriptif, di sisi lain, dimaksudkan untuk menyediakan cara untuk mengidentifikasi pola atau hubungan antara data dan memeriksa karakteristiknya[12][13].

### 2.3 K-Means

Algoritma K-Means merupakan metode non-hierarki yang pertama-tama menggunakan beberapa komponen populasi sebagai pusat cluster pertama. K-Means adalah algoritma yang digunakan dalam partisi yang membagi data ke dalam kelompok yang berbeda. Algoritma ini dapat meminimalkan jarak antara data dan cluster. Pada dasarnya, penggunaan algoritma ini dalam proses clustering tergantung pada data yang diterima dan kesimpulan yang diambil di akhir proses[14][11][15]. Oleh karena itu, saat menggunakan algoritma K-Means, aturan berikut berlaku:

- a. Jumlah cluster yang Anda butuhkan untuk memasukkan
- b. Hanya ada atribut tipe nomor

Pada dasarnya, algoritma K-Means hanya mengambil sebagian dari jumlah komponen yang diambil untuk digunakan sebagai pusat cluster awal ketika menentukan pusat cluster yang dipilih secara acak dari populasi data. Algoritma K-Means kemudian menguji setiap komponen dalam populasi data dan menandai komponen tersebut sesuai dengan jarak minimum antara komponen dan setiap pusat cluster di salah satu pusat cluster yang ditentukan sebelumnya. Selain itu, semua komponen data ditempatkan di setiap cluster dan posisi pusat cluster dihitung ulang hingga akhirnya terbentuk cluster baru.

Sedangkan menurut Theodoridis dan Koitroumbas, pada tahun , KMeans merupakan salah satu metode algoritma clustering yang paling terkenal dan populer, dan KMeans adalah prosedur literasi terbaik yang masuk dalam kategori algoritma mountaineering. Algoritma dasar untuk pengelompokan data menggunakan metode K-Means dapat dilakukan dengan cara sebagai berikut:

1. Tentukan jumlah grup
2. Secara acak berikan data ke grup
3. Hitung titik tengah (rata-rata / rata-rata) grup

Dari data yang ada di masing-masing kelompok. Lokasi centeroid setiap kelompok diambil dari rata-rata (mean) semua nilai pada setiap fitur nya. Jika  $M$  menyatakan jumlah data dalam sebuah kelompok,  $i$  menyatakan fitur ke- $i$  dalam sebuah kelompok, dan  $p$  menyatakan dimensi data, maka persamaan untuk menghitung centeroid fitur ke- $i$  digunakan persamaan 1.

$$C_i = \frac{1}{M} \sum_{j=1}^M X_j \quad (1)$$

Persamaan 1 dilakukan sebanyak  $p$ . Dimensi dari  $i=1$  sampai dengan  $i=p$

4. Alokasikan masing-masing data ke centeroid/rata-rata terdekat.

Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok. Pengukuran jarak pada ruang jarak (distance space) euclidean dapat dicari menggunakan persamaan 2.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

Pengalokasian kembali data ke dalam masing-masing kelompok dalam metode K-Means didasarkan pada perbandingan jarak antara data dengan centeroid setiap kelompok yang ada. Data dialokasikan ulang secara tegas ke kelompok yang mempunyai centeroid dengan jarak terdekat dari data tersebut. Pengalokasian data ini dapat ditentukan menggunakan persamaan sebagai berikut.

$$a_{i1} = \begin{cases} 1 & d = \min\{D(x, e_1)\} \\ 0 & \text{lainya} \end{cases} \quad (3)$$

$A_{i1}$  adalah keanggotaan titik  $x_i$  ke pusat kelompok setelah dibandingkan, dan  $c_1$  adalah centeroid (pusat kelompok) ke-1. Fungsi objektif yang digunakan untuk metode K-Means ditentukan berdasarkan jarak dan nilai ke anggotaan dalam kelompok. Fungsi objektif dapat ditentukan dengan persamaan 4.

$$\sum_{i=1}^n \sum_{j=1}^k a_{ij} D(x_i, c_j)^2 \quad (4)$$

## 3. HASIL DAN PEMBAHASAN

Pada pengklasterisasian data penyakit pasien maka digunakan data-data penyakit pasien Rumah Sakit Umum Daerah Kota Subulussalam Aceh dengan menggunakan Algoritma K-Means Clustering. Penelitian ini menggunakan beberapa materi yang digunakan: informasi pasien berdasarkan nilai gejala penyakit dan setiap tingkatannya. Fokus pada informasi sebagai fokus saat membuat grup baru. Dengan demikian, informasi yang diminta oleh setiap pasien dari sumber diagnosis penyakit dapat digabungkan dari informasi ini. Metode yang digunakan dalam penyelesaian ini adalah dengan mengumpulkan informasi, mengamati masalah, dan bereksperimen. Metode yang digunakan untuk agregasi data adalah metode clustering, yang membantu menggabungkan informasi yang kecil dan terlalu mendasar. Metode pengelompokan adalah sebagai berikut:

1. Pemilihan informasi data yang digabungkan
2. Membuat tabel database

3. Literasi 1 memilih fokus pertama. Fokus awal ditentukan secara acak dari database yang ada.
4. Jumlah jarak ke pusat cluster untuk menghitung jarak antara informasi dan pusat cluster digunakan sebagai jarak Euclidean, dan jaraknya adalah C1, C2, dan C3
5. Setelah penggabungan, penggabungan, langkah selanjutnya kembali ke langkah 3 hingga informasi penggabungan mirip dengan informasi penggabungan sebelumnya. Maka proses literasi selesai. Setelah pemaparan dilakukan, keputusan dibuat berdasarkan hasil penggabungan yang dicapai..

Dari pembahasan tersebut dijadikan sebagai contoh identitas pasien yang diinformasikan. Hasil pembahasannya adalah sebagai berikut.

1. menggunakan dataset

Pasien yang nantinya dipastikan untuk klustersasikan Jenis Penyakit Seperti tabel Berikut :

**Tabel 1.** adalah tabel dataset pasien

No	Nama	Diagnosa Penyakit	Ringan	Sedang	Berat
1	M. Dani	Tb Paru	3,54	3,7	3,54
2	Sulaiman	Tb Paru	3,27	3,4	3,18
3	Surya	Tb Paru	3,18	3,2	3
4	Sartika	Tbc	3,36	3,3	3,27
5	Dek.Maimunah	Tb Paru	3,63	3,7	3,54
6	Edy Syahputra	Tb Paru	3,27	3,5	2,91
7	Agung	Hiv Aids	3,63	3,2	3,18
8	Ruslan	Tb Paru	3,27	3,5	2,91
9	Maya Sari	Tbc	3,63	3,7	3,63
10	Al Ajimi	Hiv Aids	3,27	3,5	3,18
11	Radiah	Hiv Aids	3,45	3	3,18
12	Nita Arlaila	Hiv Aids	3,36	3,3	3
13	Masriadi	Hiv aids	3,45	3,2	3
14	Agus Syahda	Tb Paru	3,09	3,1	3
15	Jubaidah	Tbc	3,27	3,1	3,09

Penjelasan dari tabel 1 yaitu Ringan (x1) yang merupakan rata-rata nilai dari pasien yang gejala tingakatn ringan dari Tb paru, Sedang(x2) merupakan rata-rata nilai dari Pasien Yang gejala tingkatan sedang dari Tbc dan Berat (x3) merupakan rata-rata nilai dari pasien yang gejala tingkatan berat dari Hiv Aids

2. Literasi 1

Hal pertama yang dilakukan dalam literasi adalah menentukan fokus awal, yang ditentukan secara acak dari dataset yang ada. Centroid pertama diambil dari data ketiga sebagai pusat grup 1, data kedelapan sebagai pusat cluster kedua, dan data ketiga belas sebagai pusat bundel ketiga.

**Tabel 2.** Centroid Awal

No	Nama	Diagnosa Penyakit	X1	X2	X3
3	Surya	Tb Paru	3,18	3,2	3
8	Ruslan	Tb Paru	3,27	3,5	2,91
13	Masriadi	Hiv Aids	3,45	3,2	3

3. Perhitungan Jarak Pusat Cluster

Untuk mengukur jarak antara data dengan pusat Cluster digunakan *Engclidian distance*, kemudian akan didapatkan jarak yaitu C1, C2, dan C3 Sebagai berikut:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{Ki} - X_{Kj})^2} \dots (1)$$

Data ke-1 dihitung dengan data ke-1 sebagai pusat cluster 1 dengan hasil 0,81 , data ke-2 dihitung dengan data ke-1 sebagai pusat cluster 1 dengan hasil 0,28 dan data ke-3 dihitung dengan data ke-1 sebagai pusat cluster 1 dengan hasil 0 dan seterusnya.

**Tabel 3.** Perhitungan Iterasi ke-1

C	Data	Perhitungan	Hasil
1	1	$\sqrt{(3,54 - 3,18)^2 + (3,7 - 3,2)^2 + (3,54 - 3)^2}$	0,81
1	2	$\sqrt{(3,27 - 3,18)^2 + (3,4 - 3,2)^2 + (3,18 - 3)^2}$	0,28
1	3	$\sqrt{(3,18 - 3,18)^2 + (3,2 - 3,2)^2 + (3 - 3)^2}$	0

C	Data	Perhitungan	Hasil
1	4	$\sqrt{(3,38 - 3,18)^2 + (3,3 - 3,2)^2 + (3,27 - 3)^2}$	0,33
1	5	$\sqrt{(3,63 - 3,18)^2 + (3,7 - 3,2)^2 + (3,54 - 3)^2}$	0,86
1	6	$\sqrt{(3,27 - 3,18)^2 + (3,1 - 3,2)^2 + (3, - 3)^2}$	0,13
1	7	$\sqrt{(3,36 - 3,18)^2 + (3,2 - 3,2)^2 + (3,18 - 3)^2}$	0,25
1	8	$\sqrt{(3,27 - 3,18)^2 + (3,5 - 3,2)^2 + (2,91 - 3)^2}$	0,32
1	9	$\sqrt{(3,63 - 3,18)^2 + (3,7 - 3,2)^2 + (3,63 - 3)^2}$	0,92
1	10	$\sqrt{(3,27 - 3,18)^2 + (3,5 - 3,2)^2 + (3,18 - 3)^2}$	0,36
...	...	....	...
...	...	....	...
3	15	$\sqrt{(3,27 - 3,45)^2 + (3,1 - 3,2)^2 + (3,09 - 3)^2}$	0,71

Dari Perhitungan tabel diatas, Maka didapat hasil iterasi 1 pada tabel berikut ini dengan terbentuk cluster 1, cluster 2, cluster 3, dan Jarak terdekat.

**Tabel 4.** Hasil Literasi 1

No	Nama	X1	X2	X3	Cluster 1	Cluster 2	Cluster 3	T. Penyakit
1	M. Dani	3,18	3,7	3,54	0,81	0,71	0,75	0,71
2	Sulaiman	3,27	3,4	3,18	0,28	0,28	0,2	0,2
3	Surya	3,18	3,2	3	0	0,32	0,27	0
4	Sartika	3,36	3,3	3,27	0,33	0,42	0,30	0,30
5	Dek.Maimunah	3,63	3,7	3,54	0,86	0,75	0,76	0,75
6	Edy Syahputra	3,27	3,1	3	0,13	0,41	0,21	0,13
7	Agung	3,27	3,5	2,91	0,32	0	0,20	0,20
8	Ruslan	3,63	3,7	3,63	0,92	0,82	0,36	0
9	Maya Sari	3,27	3,5	3,18	0,36	0,27	0,83	0,82
10	Al Ajimi	3,45	3	3,18	0,38	0,59	0,39	0,27
11	Radiah	3,36	3,3	3	0,21	0,23	0,26	0,26
12	Nita Arlaila	3,45	3,2	3	0,27	0,36	0,13	0,13
13	Masriadi	3,09	3,1	3	0,13	0,44	0	0
14	Agus Syahda	3,27	3,1	3,09	0,16	0,43	0,34	0,13
15	Jubaidah	3,27	3,1	3,09	0,16	0,43	0,71	0,16

Jarak dari hasil perhitungan dibandingkan untuk mendapatkan jarak terdekat antara data dengan pangkal bundel. Jarak ini menunjukkan bahwa data tersebut digabungkan dengan pusat kelompok terdekat. Berikut ini adalah informasi tentang penggabungan cluster. Nilai 1 berarti informasi tersebut berada dalam satu kelompok.

**Tabel 5.** Pengelompokan Data G1

No	C1	C2	C3
1		I	
2			I
3	I		
4			I
5		I	
6	I		
7			I
8		I	
9		I	
10		I	
11			I
12			I
13			I
14	I		
15	I		



Setelah Anda mengetahui bagian-bagian untuk setiap grup, maka centroid baru dihitung berdasarkan informasi tentang bagian-bagian di setiap bundel sesuai dengan rumus dasar untuk bagian grup. Misalnya, grup pertama memiliki 5 Data, bundel kedua memiliki 4 Data, dan grup ketiga memiliki 6 Data, yang dihitung sebagai cluster baru.

**Tabel 6.** Perhitungan *Centroid* Baru

C	Perhitungan	Hasil
1	$\frac{3,18+3,09+3,27+3,18+3,36}{5}$	3,21
1	$\frac{3,2+3,1+3,1+3,7+3,3}{5}$	2,28
1	$\frac{3+3+3,09+3,54+3,27}{5}$	3,18
2	$\frac{3,27+3,63+3,36+3,45}{4}$	3,42
2	$\frac{3,4+3,7+3,3+3,2}{4}$	3,4
2	$\frac{3,18+3,54+3+3}{4}$	3,18
3	$\frac{3,27+3,36+3,27+3,63+3,27+3,45}{6}$	3,37
3	$\frac{3,1+3,2+3,5+3,7+3,5+3}{6}$	3,33
3	$\frac{3+3,18+2,91+3,63+3,18+3,18}{6}$	3,18

Centroid baru dihitung dan ditampilkan dalam tabel berikut.

**Tabel 7.** *Centroid* Baru

No	Nama	X1	X2	X3
3	Surya	3,21	3,28	3,18
8	Ruslan	3,42	3,4	3,18
13	Masriadi	3,37	3,33	3,18

4. Ulangi langkah keempat, lakukan iterasi kedua, dan gunakan rumus jarak Euclidean untuk menghitung jarak ke pusat cluster baru.

**Tabel 8.** Perhitungan Iterasi ke-2

C	Data	Perhitungan	Hasil
1	1	$\sqrt{(3,54 - 3,21)^2 + (3,7 - 3,28)^2 + (3,54 - 3,18)^2}$	0,64
1	2	$\sqrt{(3,27 - 3,21)^2 + (3,4 - 3,28)^2 + (3,18 - 3,18)^2}$	0,13
1	3	$\sqrt{(3,18 - 3,21)^2 + (3,2 - 3,28)^2 + (3 - 3,18)^2}$	0,19
1	4	$\sqrt{(3,36 - 3,21)^2 + (3,3 - 3,28)^2 + (3,27 - 3,18)^2}$	0,17
1	5	$\sqrt{(3,63 - 3,21)^2 + (3,7 - 3,28)^2 + (3,54 - 3,18)^2}$	0,69
1	6	$\sqrt{(3,27 - 3,21)^2 + (3,1 - 3,28)^2 + (3 - 3,18)^2}$	0,26
1	7	$\sqrt{(3,36 - 3,21)^2 + (3,2 - 3,28)^2 + (3,18 - 3,18)^2}$	0,17
1	8	$\sqrt{(3,27 - 3,21)^2 + (3,5 - 3,28)^2 + (2,91 - 3,18)^2}$	0,35
1	9	$\sqrt{(3,63 - 3,21)^2 + (3,7 - 3,28)^2 + (3,63 - 3,18)^2}$	0,74
1	10	$\sqrt{(3,27 - 3,21)^2 + (3,5 - 3,28)^2 + (3,18 - 3,18)^2}$	0,22
...	...	...	...
3	14	$\sqrt{(3,09 - 3,37)^2 + (3,1 - 3,33)^2 + (3 - 3,18)^2}$	0,40
3	15	$\sqrt{(3,27 - 3,37)^2 + (3,1 - 3,33)^2 + (3,09 - 3,18)^2}$	0,26

Dari hasil perhitungan iterasi kedua didapatkan hasil iterasi sebagai berikut:

**Tabel 9.** Tabel Iterasi ke-2

No	Nama	X1	X2	X3	Cluster 1	Cluster 2	Cluster 3	T. Penyakit
1	M. Dani	3,18	3,7	3,54	0,64	0,48	0,54	0,48
2	Sulaiman	3,27	3,4	3,18	0,13	0,15	0,12	0,12
3	Surya	3,18	3,2	3	0,19	0,36	0,29	0,19
4	Sartika	3,36	3,3	3,27	0,17	0,14	0,09	0,09
5	Dek.Maimunah	3,63	3,7	3,54	0,69	0,51	0,57	0,51
6	Edy Syahputra	3,27	3,1	3	0,26	0,38	0,30	0,26
7	Agung	3,27	3,5	2,91	0,17	0,20	0,13	0,13
8	Ruslan	3,63	3,7	3,63	0,35	0,32	0,33	0,32
9	Maya Sari	3,27	3,5	3,18	0,74	0,58	0,63	0,58
10	Al Ajimi	3,45	3	3,18	0,22	0,18	0,19	0,18
11	Radiah	3,36	3,3	3	0,36	0,40	0,33	0,33
12	Nita Arlaila	3,45	3,2	3	0,23	0,21	0,18	0,18
13	Masriadi	3,09	3,1	3	0,31	0,27	0,23	0,23
14	Agus Syahda	3,27	3,1	3,09	0,28	0,48	0,40	0,28
15	Jubaidah	3,27	3,1	3,09	0,21	0,34	0,26	0,21

5. Metode berikut ini sama dengan nomor 4, dimana jarak yang dihitung dibandingkan untuk menentukan jarak terdekat antara informasi dan pusat cluster. Jarak ini menunjukkan bahwa informasi tersebut berada pada kelompok dengan pusat cluster terdekat.

**Tabel 10.** Pengelompokan Data G2

No	C1	C2	C3
1		I	
2			I
3	I		
4			I
5		I	
6	I		
7			I
8		I	
9		I	
10		I	
11			I
12			I
13			I
14	I		
15	I		

G1 = G2 memiliki bagian yang tepat sehingga tidak akan berulang. Hasil clustering stabil dan konvergen.

**Tabel 11.** Tabel Penentuan Konsentrasi

Cluster	Diagnosa Penyakit	Hasil
1	Tb Paru	4
2	Tbc	5
3	Hiv Aids	6

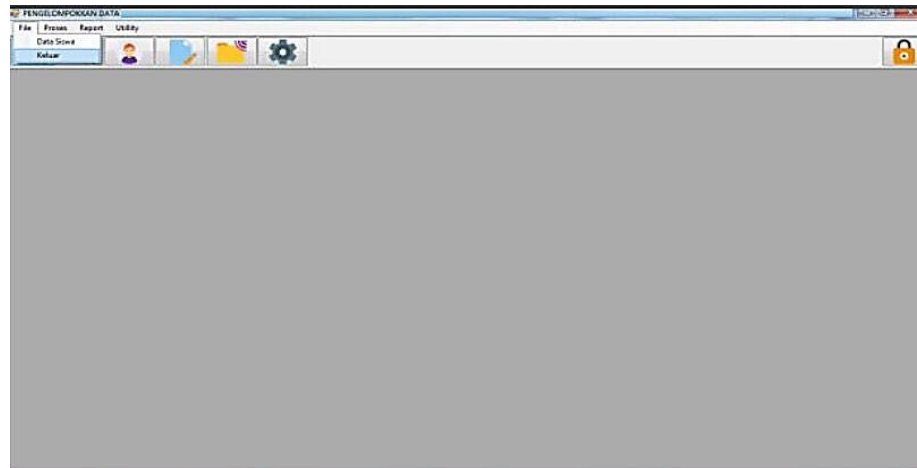
### 3.1 Hasil Pengujian

Di bawah ini adalah tampilan pengujian sistem pengelompokan data siswa menggunakan dengan metode clustering

#### 1. Menu Utama

Dibawah ini adalah tampilan antarmuka menu utama.

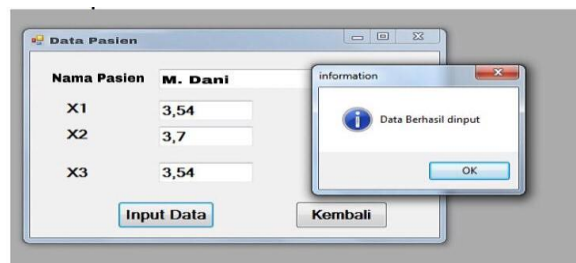




Gambar 2. Tampilan Menu Utama

## 2. Input Data Pasien

Berikut ini adalah indikator entri data pasien berdasarkan nilai tingkat diagnosis penyakit pasien.



Gambar 3. Input Data Pasien

## 3. Proses Clustering

Pengelompokan data pasien dapat dilihat pada tampilan gambar berikut ini:

Proses Clustering

Data Awal Centroid)

No	Nama	Diagnosa Penyakit	Ringan	Sedang	Berat
1	M. Dani	Tb Paru	3,54	3,7	3,54
2	Sulaiman	Tb Paru	3,27	3,4	3,18
3	Surya	Tb Paru	3,18	3,2	3
4	Sartika	Tbc	3,36	3,3	3,27
5	Dek Maimunah	Tb Paru	3,63	3,7	3,54
6	Edy Syahputra	Tb Paru	3,27	3,5	2,91
7	Agung	Hiv Aids	3,63	3,2	3,18
8	Ruslan	Tb Paru	3,27	3,5	2,91
9	Maya Sari	Tbc	3,63	3,7	3,63
10	Al Ajimi	Hiv Aids	3,27	3,5	3,18
11	Radiah	Hiv Aids	3,45	3	3,18
12	Nita Arlaila	Hiv Aids	3,36	3,3	3
13	Masriadi	Hiv aids	3,45	3,2	3
14	Agus Syahda	Tb Paru	3,09	3,1	3
15	Jubaidah	Tbc	3,27	3,1	3,09

Centroid Awal

No	Nama	Diagnosa Penyakit	X1	X2	X3
3	Surya	Tb Paru	3,18	3,2	3
8	Ruslan	Tb Paru	3,27	3,5	2,91
13	Masriadi	Hiv Aids	3,45	3,2	3

Hasil

Cluster	Diagnosa Penyakit	Hasil
1	Tb Paru	4
2	Tbc	5
3	Hiv Aids	6

Tombol

Input Data

Tampil Nilai

Proses

Hasil

Hasil Perhitungan Iterasi

No	Nama	X1	X2	X3	Cluster 1	Cluster 2	Cluster 3	T. Penyakit
1	M. Dani	3,18	3,7	3,54	0,64	0,48	0,54	0,48
2	Sulaiman	3,27	3,4	3,18	0,13	0,15	0,12	0,12
3	Surya	3,18	3,2	3	0,19	0,36	0,29	0,19
4	Sartika	3,36	3,3	3,27	0,17	0,14	0,09	0,09
5	Dek Maimunah	3,63	3,7	3,54	0,69	0,51	0,57	0,51
6	Edy Syahputra	3,27	3,1	3	0,26	0,38	0,30	0,26
7	Agung	3,27	3,5	2,91	0,17	0,20	0,13	0,13
8	Ruslan	3,63	3,7	3,63	0,35	0,32	0,33	0,32
9	Maya Sari	3,27	3,5	3,18	0,74	0,58	0,63	0,58
10	Al Ajimi	3,45	3	3,18	0,22	0,18	0,19	0,18
11	Radiah	3,36	3,3	3	0,36	0,40	0,33	0,33
12	Nita Arlaila	3,45	3,2	3	0,23	0,21	0,18	0,18
13	Masriadi	3,09	3,1	3	0,31	0,27	0,23	0,23
14	Agus Syahda	3,27	3,1	3,09	0,28	0,48	0,40	0,28
15	Jubaidah	3,27	3,1	3,09	0,21	0,34	0,26	0,21

Pengelompokan

No	C1	C2	C3
1		I	
2			I
3	I		
4			I
5		I	
6	I		
7			I
8		I	
9		I	
10		I	
11			I
12			I
13			I
14	I		
15	I		

Gambar 4. Tampilan Proses Clustering

## 4. KESIMPULAN

Berdasarkan hasil pengamatan atau penelitian yang penulis lakukan di Rumah Sakit Umum Kota Subulussalam Aceh. Maka penulisan menyimpulkan bahwa penelitian ini menggunakan data dari Rumah Sakit Umum Kota Subulussalam Aceh, sebanyak 15 sampel data pasien yang mempunyai kriteria-kriteria yang penulis butuhkan. Metode yang penulis gunakan dalam penulisan penelitian ini adalah Algoritma K-Means Clustering, dari perhitungan yang telah dilakukan, Maka direkomendasikan penulis sebanyak 3 cluster. Hasil perbandingan Perhitungan berbeda di antara cluster 1, 2, dan cluster 3 Sebagai berikut: Cluster 1 sebanyak 4 pasien, Cluster Sebanyak 5 Pasien, dan Cluster 3, Sebanyak 6 pasien. Algoritma K-Means Clustering Dapat diterapkan Pada pengklasterisasian atau penggelompokkan penyakit pasien.

## REFERENCES

- [1] S. Al Syahdan and A. Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, 2018, doi: 10.32672/jnkti.v1i2.771.
- [2] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [3] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, pp. 51–58, 2020.
- [4] D. M. Sinaga, A. P. Windarto, D. Hartama, and S. Saifullah, "Pengelompokan Indeks Harga Konsumen Menurut Kota Dengan Datamining Clustering," in *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2019, vol. 2, no. 1.
- [5] M. Anjelita, A. P. Windarto, and D. Hartama, "Pemanfaatan Datamining Pada Pengelompokan Provinsi Terhadap Pencemaran Lingkungan Hidup," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 659–666, 2019, doi: 10.30865/komik.v3i1.1675.
- [6] . F., F. T. Kesuma, and S. P. Tamba, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- [7] W. Purba, W. Siawin, and . H., "Implementasi Data Mining Untuk Pengelompokkan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019, doi: 10.34012/jusikom.v2i2.429.
- [8] "Data Mining : Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue ( Dbd ) Berdasarkan Provinsi," vol. 3, no. 2, pp. 173–178, 2018.
- [9] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. Dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [10] D. P. Indini, S. R. Siburian, N. Nurhasanah, and D. P. Utomo, "Implementasi Algoritma DBSCAN untuk Clustering Seleksi Penentuan Mahasiswa yang Berhak Menerima Beasiswa Yayasan," in *Prosiding Seminar Nasional Sosial, Humaniora, dan Teknologi*, 2022, pp. 325–331.
- [11] Y. Syahra, "Penerapan Data Mining Dalam Pengelompokkan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma K-Means Clustering," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 17, no. 2, p. 228, 2018, doi: 10.53513/jis.v17i2.70.
- [12] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth ( Fp-Growth ),," *Telematika*, vol. 4, no. 1, pp. 118–132, 2012.
- [13] B. D. Mudzakkir, "Pengelompokan Data Penjualan Produk Pada Pt Advanta Seeds Indonesia Menggunakan Metode K-Means," *J. Mhs. Tek. Inform.*, vol. 2, no. 2, pp. 34–40, 2018.
- [14] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, pp. 25–36, 2021.
- [15] S. A. Rahmah, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering ( Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja )," vol. 1, no. 1, pp. 1–5, 2020.