

## IMPLEMENTASI ALGORITMA K-MEANS UNTUK KLASTERISASI PASIEN HIPERTENSI BERSADARKAN KARAKTERISTIK PASIEN

Dani Saepul Qirom<sup>1</sup>, Ahmad Faqih<sup>2</sup>, Gifthera Dwilestari<sup>3</sup>

<sup>1,2</sup> Teknik Informatika, <sup>3</sup> Sistem Informasi, STMIK IKMI Cirebon  
Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon  
Danisaepulqirom505@gmail.com

### ABSTRAK

Dalam beberapa dekade terakhir, fokus terhadap penyakit hipertensi menjadi semakin mendesak seiring dengan perkembangan pesat di bidang Informatika yang telah memberikan dampak signifikan pada berbagai aspek kehidupan. Puskesmas yang terletak di Kecamatan Rajapolah Kabupaten Tasikmalaya mengalami peningkatan dalam jumlah pasien yang datang ke puskesmas, penyakit hipertensi paling sering dialami oleh setiap pasien yang datang ke Puskesmas Rajapolah. Penelitian ini bertujuan untuk mengelompokkan data pasien melalui pendekatan *data mining* untuk mengidentifikasi kelompok pasien hipertensi. Metode yang digunakan yaitu algoritma *k-means clustering* menggunakan software *rapidminer* dengan tahapan *Knowledge Discovery in Database (KDD)*. Dalam menentukan jumlah cluster yang paling optimal menggunakan hasil evaluasi *Davies Bouldin Index (DBI)*. Setiap 2 cluster sampai 10 cluster di evaluasi untuk mencari jumlah cluster dengan nilai DBI paling rendah. Data yang digunakan merupakan data rekam medis Puskesmas Rajapolah pada bulan september 2023 dengan jumlah 1000 records. Hasil analisis menunjukkan 4 cluster dengan nilai DBI 0.269, jumlah anggota cluster 0 : 213 pasien, cluster 1 : 302 pasien, cluster 2 : 145 pasien, cluster 3 : 68 pasien. Kelompok dengan tingkat hipertensi paling tinggi yaitu cluster 3 berjumlah 68 pasien dengan umur 30 sampai 74 tahun, tingkat hipertensi kelompok ini berada pada stadium 2 sampai krisis hipertensi.

**Kata kunci :** Hipertensi; Data mining; Clustering K-Means; Davies Bouldin Index

### 1. PENDAHULUAN

Dalam beberapa dekade terakhir, fokus terhadap penyakit hipertensi menjadi semakin mendesak seiring dengan perkembangan pesat di bidang Informatika yang telah memberikan dampak signifikan pada berbagai aspek kehidupan. Meskipun penyakit ini merupakan isu kesehatan global yang telah lama dikenal, kemajuan teknologi informasi dan komunikasi telah memperluas pemahaman kita tentang faktor-faktor risiko, metode diagnosis, dan pendekatan pengelolaan hipertensi. Hipertensi atau tekanan darah tinggi yang tidak terkontrol dapat memicu komplikasi berbahaya seperti penyakit jantung koroner, stroke, penyakit ginjal, gangguan penglihatan, bahkan kematian. Faktanya, lebih banyak orang meninggal karena tekanan darah tinggi dibandingkan penyebab lainnya.[1]

Puskesmas merupakan pusat pelayanan kesehatan yang menyediakan layanan kesehatan dasar kepada masyarakat di tingkat paling bawah, terutama di tingkat desa atau kelurahan. Puskesmas yang terletak di Kecamatan Rajapolah Kabupaten Tasikmalaya mengalami peningkatan dalam jumlah pengunjung pasien yang datang ke puskesmas, pasien dengan penyakit hipertensi paling banyak dialami oleh setiap pasien yang datang ke Puskesmas Rajapolah. Dalam konteks Informatika permasalahan yang muncul adalah *kompleksitas dan heterogenitas data* kesehatan pasien yang memerlukan pendekatan analisis yang canggih. Mengumpulkan data penyakit secara berkala meningkatkan akumulasi data [2].

Penyediaan informasi belum memenuhi kebutuhan informasi yang besar [3]. Oleh karena itu,

penelitian ini bertujuan untuk mengelompokkan pasien hipertensi dan mengidentifikasi kelompok pasien dengan tingkat hipertensi normal, prehipertensi, stadium 1, stadium 2 dan krisis hipertensi dengan menggunakan teknik *data mining*.. Untuk mencapai tujuan pekerjaan ini, saya menerapkan metode *k-means clustering* pada analisis data pasien hipertensi. Metode *k-means clustering* digunakan untuk mengelompokkan pasien berdasarkan karakteristik yang terdapat pada data. Pendekatan ini dipilih karena kemampuannya untuk mengidentifikasi kelompok homogen dalam kumpulan data yang besar, memberikan pemahaman yang lebih baik tentang variasi dan pola yang mungkin ada dalam populasi penderita hipertensi. Penelitian ini akan memanfaatkan data pasien hipertensi yang diperoleh dari Puskesmas Kecamatan Rajapolah Kabupaten Tasikmalaya. Analisis kluster menggunakan metode *k-means* akan dijalankan dengan bantuan perangkat lunak analisis data menggunakan *rapidminer*. Selanjutnya, hasil dari analisis kluster akan dianalisis secara mendalam untuk mengidentifikasi karakteristik dan pola yang muncul dalam setiap kelompok.

Beberapa penelitian terdahulu dalam konteks analisis kluster data pasien telah memberikan wawasan yang berharga, namun masih memberikan peluang untuk penelitian lebih lanjut. Kajian penerapan metode *K-means Clustering* untuk pemetaan sebaran penyakit demam berdarah. Tujuan dari penelitian ini adalah menerapkan konsep data mining menggunakan algoritma *k-means (clustering)* untuk memetakan desa-desa yang terjangkit *penyakit Demam Berdarah Dengue*. Berdasarkan hasil penelitian, terdapat 3

klaster angka prevalensi DBD yaitu. cluster 1 tingkat sedang, cluster 2 tingkat tinggi, dan cluster 3 tingkat rendah. Hasil *Devies Bouldin Index (DBI)* sebesar 1044 menunjukkan  $K = 3$  merupakan jumlah kelompok optimal dibandingkan  $K$  kelompok lain[4]. Studi kasus di Kabupaten Pinrang, Hasil penelitian pada responden Kabupaten Pinrang adalah penyebab langsung dari kejadian penyakit hipertensi (*Direct cause*) yaitu umur dan gt, 40 tahun, kebiasaan konsumsi makanan berlemak, makanan dengan kadar kolesterol tinggi garam dan faktor keturunan. Namun metode penelitian yang digunakan dalam penelitian ini adalah deskriptif dengan pendekatan penelitian kualitatif yang memanfaatkan rumusan masalah metode *root cause analysis*. [1]

## 2. TINJAUAN PUSTAKA

### 2.1. Data Mining

Penambangan data adalah penemuan pengetahuan yang tersembunyi di database [5]. *Data mining* atau pengenalan pola adalah teknik yang digunakan dalam pengolahan data untuk menemukan pola tersembunyi dalam data yang diolah. Data yang diolah dengan teknik *data mining* menghasilkan informasi baru dari data lama. Hasil pengolahan data dapat digunakan untuk pengambilan keputusan di masa depan. Istilah data mining mengacu pada cabang ilmu pengetahuan yang tujuan utamanya adalah mencari, menambang, dan mengekstrak informasi dari data yang ada[6]. Penggalan data atau data mining dapat menggunakan teknik seperti *klasifikasi, pengelompokan, regresi, pemilihan variabel, dan analisis keranjang pasar*.

### 2.2. Clustering

*Clustering* merupakan suatu metode pengelompokan berdasarkan ukuran kedekatan (kesamaan). *Clustering* berbeda dengan klasifikasi karena tidak ada variabel target untuk pengelompokan. *Clustering* tidak mencoba mengklasifikasikan, menyimpulkan, atau memprediksi nilai variabel target. Sebaliknya, algoritma pengelompokan membagi bagian dari keseluruhan kumpulan data menjadi subkelompok atau kelompok yang relatif homogen, dimana kesamaan data di dalam *cluster* dimaksimalkan dan kesamaan data di luar *cluster* diminimalkan.[2]

### 2.3. Algoritma K-Means

*K-Means* adalah teknik pengelompokan data *non-hierarki* yang mengelompokkan data ke dalam satu atau lebih klaster/grup. Algoritma ini membagi data menjadi beberapa *cluster*, data dengan karakteristik yang sama dikelompokkan ke dalam *cluster* yang sama, dan data dengan karakteristik berbeda dikelompokkan ke dalam *cluster* yang berbeda[7]. *K-Means* adalah algoritma *clustering* yang mencakup algoritma *K-Medoids* berbasis objek ditambah metode distribusi berbasis *centroid*[8]. Berikut langkah-langkah algoritma *k-means* yaitu :

- a. Tentukan jumlah *cluster* ( $k$ ) pada data set sebagai nilai *centroid*.
- b. Menghitung jarak antara data dan titik pusat *cluster* menggunakan rumus-rumus dari *Euclidian Distance*. Yang dapat dilihat pada teori jarak *Euclidian* yang dirumuskan sebagai berikut :

$$De = \sqrt{(Xi - Si)^2 + (Yi - ti)^2} \quad (1)$$

Dimana :

- $De$  adalah *Euclidean Distance*
- $i$  adalah banyaknya objek
- $(x, y)$  adalah koordinat objek
- $(s,t)$  adalah koordinat *centroid*

### 2.4. Davies Bouldin Index (DBI)

*Davis-Bouldien Index* adalah metrik untuk memeriksa/mengevaluasi hasil algoritma *clustering*. Menggunakan DBI dalam sebuah *cluster* dianggap sebagai model *clustering* yang optimal. DBI memberikan hasil yang minimal[9]. Langkah-langkah dalam menghitung nilai DBI adalah sebagai berikut :

1. *Sum of Square within-Cluster(SSW)*.

Salah satu cara untuk mengetahui keterpaduan Cluster I adalah dengan perhitungan SSW. Dengan rumus berikut :

$$SSW_i = \frac{1}{mi} \sum_{j=i}^{mi} d(X_j, C_j) \quad (2)$$

Dimana :

$mi$  = jumlah data dalam *cluster* ke- $i$

$ci$  = *centroid cluster* ke- $i$

$d(X_j, C_j)$  = jarak setiap data ke *centroid I* yang dihitung menggunakan jarak *euclidiance*.

2. *Sum Of Square Between-Cluster (SSB)*

Tujuan dari SSB adalah untuk mengetahui perbedaan atau jarak antar *cluster*. Dengan rumus sebagai berikut:

$$SSB_{ij} = d(X_i, X_j) \quad (3)$$

Dimana :

$d(X_i, X_j)$  = jarak antara data ke  $i$  dengan data ke  $j$  di *cluster* lain.

3. *Ratio (Rasio)*

Tujuan dari perhitungan *rasio (Ri,j)* adalah untuk menentukan nilai acuan antara *cluster i* dan *cluster j* serta menghitung nilai rasio pada masing-masing *cluster*. Indeks  $I$  dan  $j$  mewakili jumlah *cluster*, jika ada empat *cluster* maka akan ada empat indeks  $i, j, k$ , dan  $l$ . Tentukan nilai *rasio* menggunakan rumus berikut :

$$R_{ij, \dots, n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{i,j} + \dots + SSB_{ni,nj}} \quad (4)$$

Dimana :

$SSW_i = \text{Sum Of Square Within-Cluster}$  pada centroid  $i$

$SSB_{i,j} = \text{Sum of Square Between Cluster}$  data ke  $i$  dengan  $j$  pada cluster yang berbeda.

Pada Persamaan 2.5,  $n$  tetap merupakan jumlah cluster yang dipilih kecuali  $n_i$  sama dengan  $n_j$ .

#### 4. Davies Bouldin Index (DBI)

Menggunakan nilai rasio yang diperoleh dari Persamaan 2.5, tentukan nilai DBI dengan menggunakan perhitungan sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (Ri, j, \dots k) \quad (5)$$

Disini  $Ri,j$  adalah perbandingan nilai  $SSW$  dan  $SSB$ . Dengan menghitung Persamaan 2.6 ke Persamaan 2.5, kita menemukan bahwa  $k$  adalah jumlah cluster. Dari perhitungan *Davies Bouldin Index (DBI)* dapat disimpulkan bahwa semakin kecil nilai *Davies-Bouldin Index (DBI)* yang diperoleh (non-negatif  $\geq 0$ ), maka cluster tersebut semakin baik.

### 3. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini bertujuan untuk menggali pemahaman mendalam tentang penggunaan algoritma clustering dalam klusterisasi pasien hipertensi berdasarkan karakteristik pasien.



Gambar 1. Metode Penelitian

Deskripsi Tahapan Aktivitas :

#### 3.1. Identifikasi Masalah

- a. Menentukan Ruang Lingkup dan Tujuan  
Melakukan analisis menyeluruh terhadap situasi atau konteks yang berkaitan dengan penelitian.

Memahami faktor-faktor yang dapat mempengaruhi atau terlibat dalam permasalahan.

#### b. Analisis Situasi

Mencari informasi dari berbagai sumber, seperti literatur, laporan, data statistik, atau studi kasus yang relevan dengan topik atau masalah yang sedang diidentifikasi.

#### c. Mengumpulkan Informasi

Memahami dengan jelas ruang lingkup atau penelitian kemudian identifikasi tujuan yang ingin dicapai.

#### d. Pernyataan Masalah

Merumuskan permasalahan yang akan dipecahkan atau diteliti.

### 3.2. Studi Literature

#### a. Formulasi Permasalahan

Penyelidikan dan peninjauan terhadap literatur atau kumpulan tulisan dan publikasi ilmiah yang relevan dengan topik atau masalah penelitian.

#### b. Pencarian literature

Studi literatur dilakukan untuk memahami status pengetahuan saat ini tentang suatu topik, mengidentifikasi celah pengetahuan, dan merinci kontribusi penelitian baru yang dapat ditambahkan.

#### c. Evaluasi Data

#### d. Analisis dan Interpretasi

### 3.3. Pengumpulan Data

#### a. Kerjasama dengan Puskesmas atau sumber data relevan

Mengumpulkan data kesehatan masyarakat yang akan digunakan dalam penelitian ini. Data tersebut diperoleh melalui kerjasama dengan Puskesmas atau instansi yang memiliki akses ke data ini.

#### b. Pengumpulan data pasien, termasuk riwayat kesehatan, jenis penyakit, geolokasi, dan waktu

Data yang dikumpulkan mencakup informasi tentang pasien, termasuk riwayat kesehatan mereka, jenis penyakit yang diderita, serta atribut geografis dan temporal yang relevan.

### 3.4. Proses Tahapan KDD

#### a. Seleksi data

Data yang telah dikumpulkan akan melalui tahap preprocessing untuk memastikan kualitas dan keterpakaianya dalam analisis.

#### b. Pembersihan data

Pembersihan data akan menghilangkan data yang tidak valid atau tidak relevan.

#### c. Pengisian nilai yang hilang

Jika ada nilai yang hilang dalam data, langkah pengisian nilai yang hilang akan dilakukan.

#### d. Transformasi data jika diperlukan

Transformasi data mungkin diperlukan jika ada skala atau format yang perlu disesuaikan.

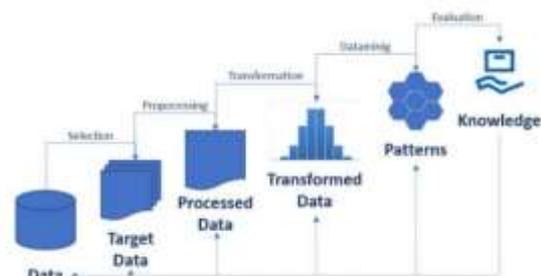
**3.5. Implementasi Algoritma K-Means**

- a. Penerapan algoritma k-means clustering pada data pasien  
Data yang telah dipreprocess akan dianalisis dengan menggunakan algoritma clustering yang sesuai. Pemilihan algoritma clustering akan bergantung pada karakteristik data dan tujuan analisis. Algoritma ini akan digunakan untuk mengelompokkan data pasien ke dalam kluster-kluster berdasarkan kemiripan karakteristik.

**3.6. Evaluasi DBI**

- a. Menentukan cluster terbaik dengan valuasi DBI  
Evaluasi Davies Bouldin Index dilakukan untuk menentukan cluster paling optimal.
- b. Visualisasi hasil clustering  
Mempresentasikan kluster-kluster yang terbentuk secara grafis. Evaluasi ini membantu dalam pemahaman awal tentang struktur data.

Metode penelitian yang menganalisis data sebagai bagian dari aplikasi *data mining* menggunakan proses tahapan *Knowledge Discovery in Databases (KDD)*. *Knowledge Discovery in Database (KDD)* merupakan langkah praktis untuk melakukan tahapan data mining untuk mencari dan mengidentifikasi pola dan data dimana pola yang ditemukan valid, berguna, dan mudah dipahami. KDD memiliki beberapa fase yang terdiri dari data, seleksi dan pembersihan data, transformasi data, penambangan data, pengembangan pola, dan pengetahuan. Di bawah ini adalah diagram setiap fase KDD :



Gambar 2. Knowledge Discovery in Database (KDD)

Adapun deskripsi pada tahapan *Knowledge Discovery in Database (KDD)* yaitu :

1. **Data**  
Data merupakan kumpulan informasi operasional yang diperlukan sebelum memulai tahap *Knowledge Discovery Database (KDD)*.
2. **Data selection**  
Proses pemilihan data (*data selection*) dalam KDD melibatkan identifikasi dan ekstraksi subset data yang relevan dari sumber data yang lebih besar. Tujuan utamanya adalah untuk fokus pada bagian data yang paling mungkin mengandung pola, informasi, atau relasi yang signifikan.
3. **Pre-processing**  
Langkah ini digunakan untuk menambah data, menghapus data, dan menghasilkan *noise*. Data

kosong atau nilai yang hilang dihilangkan dalam penelitian ini.

4. **Transformation**

Proses *transformasi data* merupakan suatu proses pengolahan data yang dapat diolah menjadi suatu format yang dapat diproses dengan proses data mining. Transformasi dapat dilakukan untuk memperbaiki distribusi data, mengurangi dimensi, atau memudahkan pemahaman pola yang terkandung dalam data. *Transformasi data* dilakukan dengan menginisialisasi data yang bernilai nominal menjadi nilai numerik.

5. **Data mining**

Pada tahap ini menggunakan algoritma atau metode pengambilan informasi. Ini adalah langkah penting dalam menggunakan teknik intelijen untuk mengekstrak pola data yang berpotensi berguna dari data yang dipilih.

6. **Evaluation**

Tahap evaluasi mengungkapkan apakah hasil tahap data mining dapat mencapai tujuan yang telah ditetapkan. Oleh karena itu, untuk setiap *cluster* yang terbentuk dibuat profil untuk mengetahui karakteristik kelompoknya.

**4. HASIL DAN PEMBAHASAN**

**4.1. Dataset**

Data pasien yang digunakan adalah data rekam medis. Rekam medis atau *ICD (International Classification of Diseases)* dalam bidang medis, adalah catatan pasien yang dirawat di rumah sakit atau klinik[10]. Data diambil langsung dari Puskesmas Rajapolah dari tanggal 1 sampai 30 September 2023 dengan jumlah 1000 recods dan 82 atribut. atribut terdiri dari No, Tanggal, Nama Pasien, No. eRM, NIK, No. KK, No. RM Lama, No. Dokumen RM, Jenis Kelamin, No Telp, Alamat, RT, RW, Pekerjaan, Tanggal Pemeriksaan, Kelurahan, Tempat Lahir, Tgl.Lahir, Umur Tahun, Umur Bulan, Umur Hari, Nama Ayah, Nama Ibu, Jenis Kunjungan, Poli/Ruangan, Asuransi, No. Asuransi, Kelainan, Dokter / Tenaga Medis, Perawat / Bidan / Nutritionist / Sanitarian, SOAP Assessment, SOAP Subjective, SOAP Objective, SOAP Planning, Keluhan Utama, Keluhan Tambahan, Lama Sakit, Merokok, Konsumsi Alkoho, Kurang Sayur/Buah, Terapi, Edukasi, Tindakan Keperawatan, Keterangan, RPS, RPD, RPK, Alergi, Kesadaran, Triage, Tinggi, Berat Badan, Lingkar Perut, IMT, Hasil IMT, Sistole, Diastole, Nafas, Detak Nadi, Detak Jantung, Suhu, Aktifitas Fisik dan Assessment Fungsional, Skala Nyeri, Diagnosa 1, Jenis Kasus 1, Diagnosa 2, Jenis Kasus 2, Diagnosa 3, Jenis Kasus 3, Diagnosa 4, Jenis Kasus 4, Diagnosa 5, Jenis Kasus 5, Tindakan, Resep, Apoteker, Pendaftaran/Rujukan Internal, Lama Antrean, Lama Pemeriksaan, Lama Pelayanan Obat, Petugas Pendaftaran.

### 4.2. Data Selection

Proses pemilihan data (*data selection*) dalam KDD melibatkan identifikasi dan ekstraksi subset data yang relevan dari sumber data yang lebih besar. Tujuan utamanya adalah untuk fokus pada bagian data yang paling mungkin mengandung pola, informasi, atau relasi yang signifikan. Atribut yang digunakan adalah atribut nama pasien, umur tahun, sistole, diastole. Sebelum tahap *pre-processing* data dengan software *rapidminer* data dengan jenis nominal dirubah terlebih dahulu kedalam numerical dengan MS.Excel . Berikut gambar data hasil selection dan transformation :

1	Nama Pasien	Umur Tahun	Sistole	Diastole
2	LUTFI ASHARI	20	111	82
3	CICI CAHYATI	59	132	74
4	DEDEH JUBAIDAH	63		
5	REDI	62		
6	ATANG ENJANG	45	100	80
7	TIAH S	61	110	70
8	N. IA MARLINA	55	130	80
9	RINA	46	141	80
10	TUTI	63		
11	RISHA ARSITA	26	102	69
12	AI DEA SYADRAH	23	120	80
13	EMAN	70		
14	INTAN ZAINA FAUZIAH	7	90	60
15	KIKI AMINAH	30		
16	KIKI AMINAH	30	115	74
17	YOYO SUNARYO	45	120	80
18	NUJUN	15	120	74
19	FATHAN ARSANA MAULANA	6	100	70
20	NUJLAILA	12	108	74

Gambar 3. Hasil selection dan transformation

### 4.3. Pre-processing

Langkah *pre-processing* digunakan untuk augmentasi data, penghapusan data, dan *noise* dengan menggunakan perangkat lunak bernama *Rapidminer*. Data kosong atau nilai yang hilang akan dihapus. Kolom kosong atau nilai yang hilang dan baris yang identik dihilangkan untuk menghindari kesalahan pada langkah berikutnya. Berikut gambar yang sudah dilakukan *Pre-processing*.

Row No.	Nama Pasien	Sistole	Diastole	Umur Tahun
1	LUTFI ASHARI	111	82	20
2	CICI CAHYATI	132	74	59
3	DEDEH JUB...	121	79	63
4	REDI	121	79	62
5	ATANG ENJA...	100	80	45
6	TIAH S	110	70	61
7	N. IA MARLINA	130	80	55
8	RINA	141	80	46
9	TUTI	121	79	63
10	RISHA ARSITA	102	69	26
11	AI DEA SYAD...	120	80	23

ExampleSet (728 examples, 1 special attribute, 3 regular attributes)

Gambar 4. Hasil pre-processing

Hasil dari *pre-processing* data menjadi 728 records.

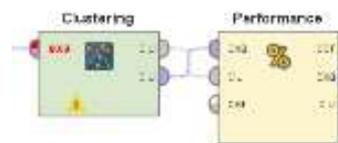
### 4.4. Data Mining

Algoritma yang digunakan adalah *k-means clustering*. Operator yang digunakan yaitu *k-means*.

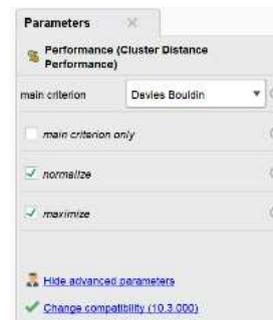


Gunakan hasil evaluasi DBI untuk menentukan ukuran kelompok yang optimal. *Davis Bouldien Index (DBI)* adalah metrik untuk memeriksa/mengevaluasi hasil algoritma *clustering*. Menggunakan DBI dalam sebuah *cluster* dianggap sebagai model *clustering* yang optimal. DBI memberikan hasil yang minimal[9]. Langkah untuk menentukan cluster terbaik dengan hasil evaluasi DBI :

1. Menambahkan operator *cluster distance performance*



2. Rubah *main criterion* atau kriteria utama menjadi *Davies Bouldin*



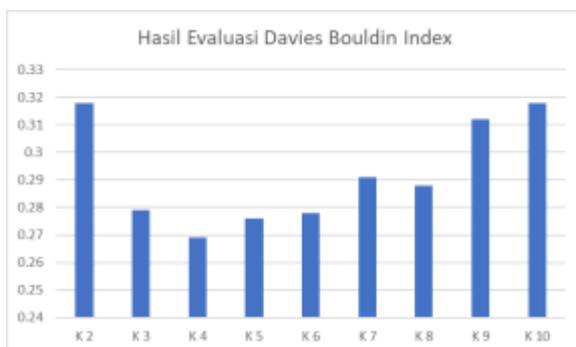
3. Evaluasi setiap cluster dari cluster 2 sampai 10.

Table 1. Hasil Evaluasi DBI

Cluster Set	Jumlah Anggota Cluster	Hasil Nilai DBI
2	Cluster 0 : 410 items Cluster 1 : 318 items	0.318
3	Cluster 0 : 244 items Cluster 1 : 76 items Cluster 2 : 408 items	0.279
4	<b>Cluster 0 : 213 items</b> <b>Cluster 1 : 302 items</b> <b>Cluster 2 : 145 items</b> <b>Cluster 3 : 68 items</b>	<b>0.269</b>
5	Cluster 0 : 190 items Cluster 1 : 294 items Cluster 2 : 144 items Cluster 3 : 19 items Cluster 4 : 81 items	0.276

Cluster Set	Jumlah Anggota Cluster	Hasil Nilai DBI
6	Cluster 0 : 127 items Cluster 1 : 123 items Cluster 2 : 244 items Cluster 3 : 80 items Cluster 4 : 135 items Cluster 5 : 19 items	0.278
7	Cluster 0 : 132 items Cluster 1 : 218 items Cluster 2 : 122 items Cluster 3 : 66 items Cluster 4 : 99 items Cluster 5 : 19 items Cluster 6 : 72 items	0.291
8	Cluster 0 : 97 items Cluster 1 : 215 items Cluster 2 : 144 items Cluster 3 : 112 items Cluster 4 : 63 items Cluster 5 : 69 items Cluster 6 : 19 items Cluster 7 : 39 items	0.288
9	Cluster 0 : 31 items Cluster 1 : 15 items Cluster 2 : 106 items Cluster 3 : 67 items Cluster 4 : 144 items Cluster 5 : 185 items Cluster 6 : 41 items Cluster 7 : 107 items Cluster 8 : 62 items	0.312
10	Cluster 0 : 52 items Cluster 1 : 160 items Cluster 2 : 112 items Cluster 3 : 103 items Cluster 4 : 43 items Cluster 5 : 106 items Cluster 6 : 15 items Cluster 7 : 31 items Cluster 8 : 35 items Cluster 9 : 71 items	0.318

Hasil gambar evaluasi DBI dari k 2 sampai 10 menunjukkan :



Gambar 5. Hasil evaluasi DBI

Hasil analisis dalam menentukan cluster terbaik dengan evaluasi DBI menunjukkan bahwa 4 cluster lebih baik dibandingkan jumlah cluster lainnya karena menghasilkan nilai yang paling rendah atau mendekati

0, dengan hasil nilai 0.269. Hasil nilai centroid yang didapatkan dari masing-masing atribut :

Cluster	Nama	Diastole	Umur Tahun
Cluster 0	123.984	70.440	17.240
Cluster 1	127.980	80.900	16.070
Cluster 2	96.810	64.607	18.880
Cluster 3	185.770	87.350	17.870

Gambar 6. Nilai Centroid dari masing-masing atribut

Hasil dataset yang diperoleh :

Table 2. Cluster\_0

No	Nama Pasien	Umur Tahun	Sistole	Diastole
1	CICI CAHYATI	59	132	74
2	DEDEH JUBAEDAH	63	121	79
3	REDI	62	121	79
4	ATANG ENJANG	45	100	80
5	TIAH S	61	110	70
6	N. IA MARLINA	55	130	80
7	RINA	46	141	80
8	TUTI	63	121	79
9	EMAN	70	121	79
10	YOYO SUNARYO	45	120	80
...	.....	.....	.....	.....
.				

Table 3. Cluster\_1

No	Nama Pasien	Umur Tahun	Sistole	Diastole
1	AI DEA SYADIAH	23	120	80
2	KIKI AMINAH	30	121	79
3	NURIN	13	126	74
4	AULIA SABILA	15	122	104
5	SINTA PRATAMA	15	122	87
6	TAY SRI RAHYU	26	110	70
7	ABU LULU MUHAAD HAQ	16	110	80
8	DADANG HENDRA LESMANA	34	120	80
9	ANTI ANDINI	20	132	72
10	ANEU NURAENI	19	120	80
...	.....	.....	.....	.....
.				

Table 4. Cluster\_2

No	Nama Pasien	Umur Tahun	Sistole	Diastole
1	LUTFI ASH'ARI	20	111	62
2	RISHA ARSITA	26	102	69
3	INTAN ZAINA FAUZIAH	7	90	60
4	FATHAN ARSAKA MAULANA	6	100	70
5	NURLAILA	12	108	74
6	LYRA	15	109	70
7	AZZAM ZAIDAN SATHIR	7	80	50
8	HISHAM HAZAMI AWWAB	12	88	60

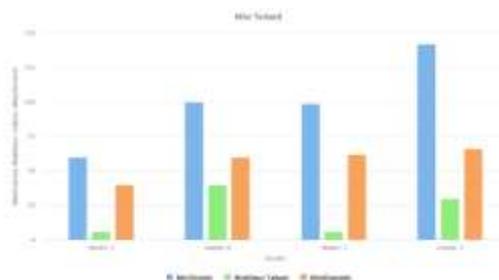
No	Nama Pasien	Umur Tahun	Sistole	Diastole
9	KEYSHA KHAIRUNISA	10	90	60
10	MUH ATALA	6	110	78
...	.....	....	....	....

Table 5. Cluster\_3

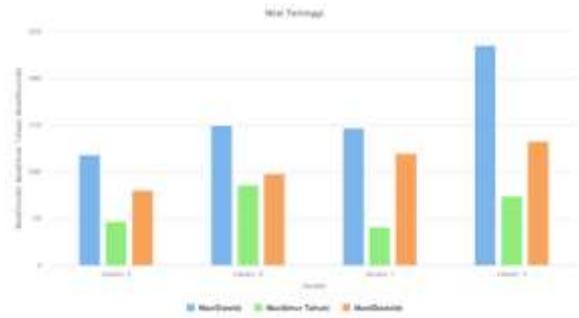
No	Nama Pasien	Umur Tahun	Sistole	Diastole
1	ONAH	46	152	87
2	LILIS FARIDA	49	182	102
3	JOJOH	59	189	104
4	YATI ROHAYATI	66	150	90
5	ATON PATONAH	47	192	122
6	KONAH	57	190	73
7	NANI SUMARNI	40	183	111
8	AAS MULYANI	33	160	106
9	TATANG KURNIAWAN	41	167	106
10	KINKIN	58	202	114
...	.....	....	....	....

Pada tabel 2-5 diatas merupakan daftar nama-nama pasien yang sudah di cluster dengan atribut yang berkaitan dengan penyakit hipertensi yaitu umur tahun, sistole(tekanan darah pada saat jantung memompa darah kedalam pembuluh nadi), diastole(tekanan darah pada saat jantung mengembang dan menyedot darah kembali) dimana cluster\_0 : 213 pasien, cluster\_1 : 302 pasien, cluster\_2 : 145 pasien, cluster\_3 : 68 pasien.

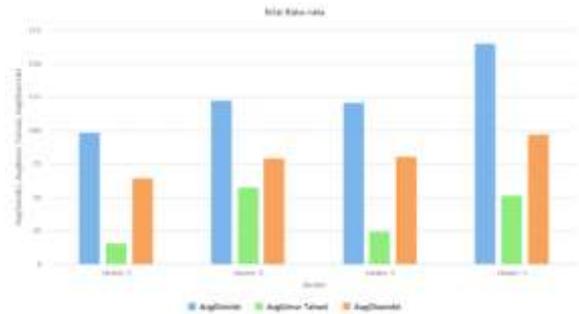
Menganalisis cluster diperlukan untuk mengidentifikasi karakter-karakter yang ada pada setiap cluster. Cluster yang memiliki anggota paling banyak adalah cluster\_1 yang berjumlah 302 pasien, cluster\_0 : 213 pasien, cluster\_2 : 145 pasien, cluster\_3 : 68 pasien.



Gambar 7. Cluster dengan atribut nilai terendah



Gambar 8. Cluster dengan nilai atribut tertinggi



Gambar 9. Cluster dengan nilai atribut rata-rata

Dari gambar 7, 8, 9 bisa di jelaskan bahwa :

- Pada cluster\_0 Berjumlah 213 pasien dengan rentang usia 40 – 86 tahun, sistole 100 mm – 150 mm, diastole 60 Hg – 98 Hg. Tingkat hipertensi pada kelompok ini rata-rata prehipertensi dan hanya beberapa yang mengalami hipertensi stadium 1 sampai stadium 2.
- Pada cluster\_1 Berjumlah 302 pasien dengan rentang usia 6 – 41 tahun, sistole 99 mm – 147 mm, diastole 62 Hg – 120 Hg. Tingkat hipertensi pada kelompok ini rata-rata prehipertensi dan hanya beberapa yang mengalami hipertensi stadium 1 sampai stadium 2.
- Pada cluster\_2 berjumlah 145 pasien dengan umur 6 – 47 tahun, sistole 60 mm – 118 mm, diastole 40 Hg – 80 Hg. Kelompok ini berada pada tingkat normal.
- Pada Cluster\_3 berjumlah 68 pasein dengan umur 30 - 74 tahun, sistole 142 mm - 235 mm, diastole 66 Hg - 133 Hg. Kelompok berada pada tingkat stadium 2 dan krisis hipertensi.

Cluster dengan tingkat hipertensi paling tinggi yaitu pada cluster\_3 yang berjumlah 68 pasein dengan umur 30 - 74 tahun, sistole 142 mm - 235 mm, diastole 66 Hg - 133 Hg, tingkat hipertensi kelompok ini berada pada stadium 2 sampai krisis hipertensi, dan cluster dengan tingkat hipertensi paling rendah atau normal yaitu pada cluster\_2 yang berjumlah 145 pasien dengan umur 6 - 47 tahun, sistole 60 mm - 118 mm, diastole 40 Hg – 80 Hg. Berdasarkan analisis bahwa pasien yang mengalami hipertensi di puskesmas Rajapolah pada tanggal 1 sampai 30 september 2023 rata-rata dialami oleh pasien dewasa sampai lansia.

## 5. KESIMPULAN DAN SARAN

Implementasi algoritma *k-means clustering* pada data pasien digunakan bantuan software *rapidminer* sebagai alat untuk menganalisis data dan tahapan KDD sebagai teknik analisis data. Data penelitian ini bersumber dari puskesmas yang beralamat di Jl. Raya Rajapolah, Kec. Rajapolah, Kab. Tasikmalaya, 46155. Data yang digunakan adalah data rekam medis Selama satu bulan dari tanggal 1 September sampai dengan 30 September 2023 berjumlah 1000 records dan 82 atribut.

Perhitungan *Davies-Bouldin Index (DBI)* menunjukkan bahwa cluster 4 mempunyai nilai k terbaik, dengan nilai terendah atau mendekati 0 pada percobaan untuk 2 sampai 10 cluster dengan nilai 0.269, jumlah anggota cluster 0 : 213 pasien, cluster 1 : 302 pasien, cluster 2 : 145 pasien, cluster 3 : 68 pasien. Kelompok dengan tingkat hipertensi paling tinggi yaitu pada cluster 3 yang berjumlah 68 pasien dengan umur 30 sampai 74 tahun, tingkat hipertensi kelompok ini berada pada stadium 2 sampai krisis hipertensi. Pasien yang mengalami hipertensi di puskesmas Rajapolah pada tanggal 1 sampai 30 september 2023 rata-rata dialami oleh pasien dewasa sampai lansia.

Oleh karena itu saran yang dapat diberikan dari penelitian ini adalah petugas kesehatan untuk lebih mengedepankan dan memperhatikan waktu tunggu pasien lanjut usia yang terjangkit hipertensi dan sebaiknya para profesional kesehatan bekerjasama dengan masyarakat untuk memberikan penyuluhan tentang penyakit hipertensi terutama masyarakat yang sudah dewasa dan lanjut usia dalam pencegahan dan pengendalian penyakit hipertensi.

## DAFTAR PUSTAKA

- [1] M. Maulia, H. K. Hengky, dan H. Muin, "Analisis Kejadian Penyakit Hipertensi di Kabupaten Pinrang," *J. Ilm. Mns. dan Kesehat.*, vol. 4, no. 3, hal. 324–331, 2021.
- [2] F. Rahmadayanti, I. Anggraini, dan T. Susanti, "Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan," *J. Inf. Syst. Res.*, vol. 4, no. 2, hal. 737–741, 2023, doi: 10.47065/josh.v4i2.2905.
- [3] A. Elang *et al.*, "Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 3, hal. 172–182, 2023.
- [4] M. A. Sembiring, R. Tama, dan A. Agus, "Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue ( DBD )," *J. Sci. Soc. Res.*, vol. 4307, no. 3, hal. 336–341, 2021.
- [5] N. Agung, N. Odi, dan D. Gifthera, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, hal. 1–7, 2022.
- [6] I. N. M. Adiputra, "Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means," *Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, hal. 99–105, 2021.
- [7] R. Anggraini, E. Haerani, J. Jasril, dan I. Afrianty, "Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 6, hal. 1840, 2022, doi: 10.30865/jurikom.v9i6.5145.
- [8] T. Tendean, W. Purba, dan M. Kom, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means," *SAINTEK (Jurnal Sains dan Teknol.)*, vol. 1, no. 2, hal. 5–11, 2020.
- [9] A. Febrian, S. Nana, dan D. Gifthera, "Penerapan Algoritma K-Means Untuk Mengelompokkan Data Pengiriman Paket Di Kantor Pos Cirebon," *J. Teknol. Technoscientia*, vol. 15, no. 1, hal. 23–27, 2022, doi: 10.34151/technoscientia.v15i1.3858.
- [10] R. Ordila, R. Wahyuni, Y. Irawan, dan M. Y. Sari, "Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering ( Studi Kasus : Poli Klinik Pt . Inecda )," *J. Ilmu Komput.*, vol. 9, no. 2, hal. 148–153, 2020.