

Klasifikasi Penyakit Hipertensi Menggunakan Metode K-Means Clustering

Ummi Qalsum

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar

Wahyuni Abidin

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, wahyuniabidin@uin-alauddin.ac.id

ABSTRAK, Tekanan darah merupakan faktor yang penting dalam sirkulasi dalam tubuh, tinggi atau rendah tekanan darah dapat mempengaruhi homeostasis dalam tubuh. Dalam artikel ini bertujuan mengklasifikasikan penyakit hipertensi di Puskesmas Jongaya kota Makassar menggunakan metode K-Means. Hasil analisis menunjukkan bahwa penderita hipertensi dipuskesmas Jongaya diklasifikasikan menjadi dua kelompok yakni pasien dengan tingkat keparahan tinggi dan pasien dengan tingkat keparahan yang rendah. Berdasarkan hasil analisis, diperoleh 69 pasien dikategorikan pada tingkat keparahan rendah dan 144 pasien dikategorikan pada tingkat keparahan tinggi. Atribut yang digunakan adalah usia, sistolik, diastolik dan berat badan.

Kata Kunci: Hipertensi, data mining, k-means clustering

1. PENDAHULUAN

Tekanan darah merupakan faktor yang penting dalam sirkulasi dalam tubuh, tinggi atau rendah tekanan darah dapat mempengaruhi homeostasis dalam tubuh. Terdapat dua kelainan tekanan darah di dalam tubuh, antara lain yang di kenal sebagai hipertensi atau tekanan darah tinggi dan hipotensi atau tekanan darah rendah. Hipertensi adalah peningkatan tekanan darah $>140/90$ mmHg [1].

Hipertensi merupakan penyebab kematian nomor satu di dunia dan hipertensi menjadi penyebab kematian nomor tiga setelah stroke dan tuberculosis. Hipertensi sudah menjadi masalah kesehatan masyarakat (*Public Health Problem*) dan akan menjadi masalah yang lebih besar jika tidak ditanggulangi sejak dini [2].

Di Asia, tercatat 38,4 juta penderita hipertensi pada tahun 2000 dan diprediksi menjadi 67,4 juta orang pada tahun 2025. Sedangkan Asia Tenggara sendiri berada di posisi ke-3 tertinggi dengan prevalensi sebesar 25% terhadap total penduduk. Di Indonesia, mencapai 17,21% dari populasi penduduk dan kebanyakan tidak terdeteksi [3].

Berdasarkan Data Dinas Kesehatan Provinsi Sulawesi Selatan tahun 2020, menurut data Kabupaten/Kota prevalensi hipertensi

tertinggi terdapat di Kota Makassar sebanyak 290.247 kasus. Tingginya prevalensi hipertensi menyebabkan sering disebut sebagai penyakit tidak menular paling mematikan atau biasa disebut *The Silent Killer* karena sering kali dijumpai tanpa gejala.

Puskesmas Jongaya merupakan salah satu Puskesmas yang ada di kota Makassar, Kecamatan Tamalate Kota Makassar. Dengan peningkatan kejadian hipertensi yang mempengaruhi kesehatan masyarakat di Puskesmas Jongaya, dan jumlah peserta hipertensi dewasa pada tahun 2017 ada sebanyak 576 dan mengalami peningkatan di tahun 2018 sebanyak 580 peserta pada penderita hipertensi dewasa.

Sistem klasifikasi merupakan salah satu sistem yang dapat membantu dokter dalam mengklasifikasikan penyakit hipertensi pasien. Tujuannya adalah untuk mengetahui tingkat keparahan hipertensi yang diderita pasien dengan mengambil informasi yang diperoleh menggunakan model-model klasifikasi. K-Means merupakan salah satu metode data clustering non hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok.

Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster atau kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster atau kelompok yang lain sehingga data yang berada dalam satu cluster atau kelompok memiliki tingkat variasi yang kecil.

2. TINJAUAN PUSTAKA

Hipertensi

Hipertensi adalah keadaan seseorang mengalami peningkatan tekanan darah di atas nilai normal yang dapat menyebabkan rasa sakit hingga kematian. Tekanan darah normal berada pada 120/80 mmHg. Batasan nilai tekanan darah dikenal dengan ketetapan JNC VII (*The Seventh*

Report of The Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure) [4].

Penderita hipertensi berasal dari segala umur dan kelas kelompok ekonomi. Penyebab hipertensi disebabkan oleh dua faktor, yaitu faktor keturunan dan pola hidup. Pengobatan hipertensi bersifat jangka panjang dan seumur hidup. Pendeteksian penyakit hipertensi selama ini hanya dilakukan secara manual [5].

Data Mining

Data mining adalah sebuah tahapan semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang bermanfaat yang tersimpan didalam database yang besar [6]. Data mining juga mempunyai peran yang sangat penting dalam beberapa bidang dalam kehidupan sehari-hari antara lain ialah bidang industri, bidang cuaca, keuangan dan bidang ilmu dan teknologi. Data mining dapat dikelompokkan menjadi dua kategori utama, yaitu:

- a) *Descriptive mining*, ialah proses untuk menemukan karakteristik penting dari data dalam suatu database. Teknik data mining yang termasuk *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
- b) *Predictive*, yaitu suatu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa mendatang. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi.

K-Means

K-Means merupakan salah satu metode pengelompokkan data non-hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam kelompok yang lain [7]. Secara umum, cara kerja dari algoritma K-Means clustering adalah sebagai berikut:

- a) Menentukan *k* sebagai jumlah kluster yang diinginkan.

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster yang akan membentuk siku pada suatu titik. *Sum of Square Error* (SSE) sering digunakan dalam acuan penelitian yang berhubungan dengan optimalisasi kluster. Akan ada kluster yang mengalami penurunan paling ekstrim yang dilanjutkan dengan nilai *K* yang akan turun secara perlahan-lahan sampai nilai *K* tersebut stabil dengan persamaan pada rumus 2.1.

$$SSE = \sum_{i=1}^k \sum_{x_i \in C_i} D(x_i, c_i)^2 \quad (2.1)$$

dimana,

- k* = Jumlah kluster
- $x_i \in C_i$ = Nilai keanggotaan titik data x_i ke pusat kelompok C_i
- C_i = Pusat kluster ke-*i* D
- (x_i, C_i) = Jarak dari titik x_i ke kelompok C_i yang diikuti

- b) Menentukan nilai random untuk pusat centroid (cluster awal) sebanyak *k*.
- c) Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidean (*Euclidean Distance*) sampai menemukan jarak terdekat dari setiap data dengan centroid. Berikut adalah persamaan pada rumus:
Euclidean Distance :z

$$D = |x - y|_C = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2)$$

dimana,

- x* = data kriteria,
- y* = centroid pada cluster ke-*j*,
- d) Mengklasifikasikan data berdasarkan kedekatannya dengan centroid (jarak terkecil).
- e) Memperbaharui nilai centroid. Nilai centroid baru diperoleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus:

$$b_j(t + 1) = \frac{1}{NS_j} \sum_{j \in S_j} a_j \quad (2.3)$$

dimana,

$$b_{j(t+1)} = \text{centroid baru pada iterasi ke } (t+1)$$

$$NS_j = \text{banyak data pada cluster } S_j$$

- f) Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap cluster tidak ada yang berubah.

3. METODOLOGI

Data rekam medis dari salah satu puskesmas di kota Makassar rentang waktu 1-31 Agustus 2023 menjadi obyek dari klasifikasi k -means yang berkaitan dengan penyakit hipertensi. Variabel yang diperhatikan adalah usia, tekanan darah sisolik dan diastolsik, serta berat badan.

Prosedur Analisis

Langkah-langkah klasifikasi k -means pasien hipertensi, dijelaskan sebagai berikut:

1. Mengimport data
2. Mencari k optimum.
3. Menghitung pusat awal cluster secara acak.
4. Menghitung jarak pusat setiap cluster
5. Mengklasifikasikan data berdasarkan kedekatannya dengan centroid.
6. Memperbaharui nilai centroid.
7. Melakukan perulangan dari langkah 5 hingga 7, sampai anggota tiap cluster tidak ada yang berubah.
8. Melakukan clustering
9. Membuat plot

4. HASIL DAN PEMBAHASAN

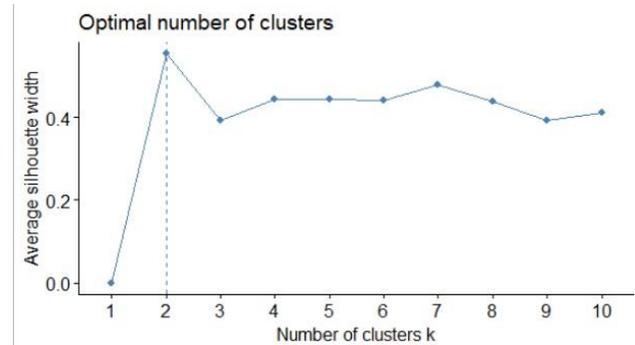
Import data

Adapun data yang digunakan dalam penelitian ini bersumber dari puskesmas Jongaya kota Makassar dan dapat dilihat pada lampiran.

Mencari K-optimum

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan

nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik. Adapun grafik yang didapatkan sebagai berikut:



Gambar 4.1 Grafik jumlah cluster

Dari Gambar 1 diatas dapat disimpulkan bahwa cluster yang akan dibentuk sejumlah 2.

Menghitung pusat awal cluster secara acak

Menghitung jarak setiap data input terhadap masing-masing centroid yang didapatkan secara random, untuk penentuan awal cluster adalah:

Pusat cluster 1: (50,167,88,59)

Pusat cluster 2: (18,30,122,81,49)

Menghitung jarak pusat setiap cluster

Untuk menghitung jarak antara data dengan pusat cluster digunakan *Euclidion distance*, kemudian akan didapatkan matrik jarak sebagai berikut:

$$D = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dari 213 data yang dijadikan sampel telah dipilih pusat awal cluster yaitu C1:(50,167,88,59), C2:(18,30,122,81,49). Kemudian dijadikan perhitungan jarak dari sisa sampel data dengan pusat cluster.

Mengklasifikasikan data berdasarkan kedekatannya dengan centroid

Cara mengklasifikasikan nilai data kedalam cluster pertama dan kedua adalah dengan membandingkan nilai cluster pertama dengan cluster kedua. Apabila C1 lebih kecil dari C2 maka pasien dimasukkan dalam cluster pertama. Dari perhitungan yang dilakukan Jumlah pasien yang masuk pada cluster pertama adalah 151 sedangkan cluster kedua berjumlah 62 pasien.

Memperbarui nilai centroid

Tentukan posisi baru (C_k) dengan cara menghitung rata-rata dari data-data yang ada pada centroid yang sama.

$$C_k = \frac{1}{n_k} \sum d_1$$

Dimana n_k adalah jumlah dokumen dalam cluster k dan d_1 adalah dokumen dalam cluster. Sehingga didapatkan nilai centroid baru yakni:

Pusat cluster 1: (60,159,90,63)

Pusat cluster 2: (69,128,78,62)

Melakukan perulangan dari langkah 5 hingga 7, sampai anggota tiap cluster tidak ada yang berubah.

Dari penentuan centroid yang dilakukan terdapat 11 kali perulangan yang sampai jumlah pasien didalam cluster tidak berubah. Nilai setiap centroid dapat dilihat pada Tabel 1.

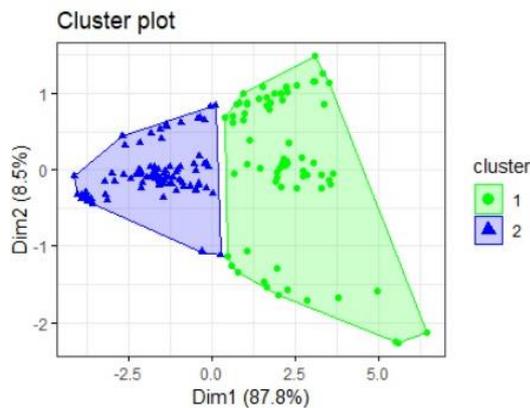
Tabel 4.1 Nilai Setiap Centroid.

Iterasi	Data Ke	Usia	Sistolik	Diastolsik	Bb
Iterasi 1	4	50	167	88	59
	18	30	122	81	49
Iterasi 2		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	60	159,2251656	89,95364238	63,45033113
	Cluster 2	59,93548387	124,3387097	77,96774194	77,96774194
Iterasi 3		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	58,75	161,9393939	91,41666667	60,67424242
	Cluster 2	61,98765432	128,0987654	78,39506173	62,2962963
Iterasi 4		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	57,8490566	167,3207547	95,16037736	64,11320755
	Cluster 2	62,09345794	130,9906542	77,85046729	58,4953271
Iterasi 5		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	57,23863636	171,4318182	97,27272727	66,15909091
	Cluster 2	61,912	133,328	78,856	57,864
Iterasi 6		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	57,29487179	174,4102564	98,44871795	66,30769231
	Cluster 2	61,53333333	134,4296296	79,54074074	58,39259259
Iterasi 7		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	56,82191781	175,6438356	99,54794521	66,60273973
	Cluster 2	61,62857143	135,2142857	79,64285714	58,52142857
Iterasi 8		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	56,50704225	176,2394366	99,83098592	66,5915493
	Cluster 2	61,71830986	135,4859155	79,78169014	58,64084507
Iterasi 9		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	56,54285714	176,6142857	99,97142857	66,28571429
	Cluster 2	61,66433566	135,5874126	79,85314685	58,84615385
Iterasi 10		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	56,56521739	176,9710145	100,0869565	66,15942029
	Cluster 2	61,61805556	135,7013889	79,9375	58,95833333
Iterasi 11		Usia	Sistolik	Diastolsik	Bb
	Cluster 1	56,56521739	176,9710145	100,0869565	66,15942029
	Cluster 2	61,61805556	135,7013889	79,9375	58,95833333

Melakukan clustering

Karena telah mendapatkan k -optimim maka yang dilakukan selanjutnya adalah pengelompokan pasien yang terkena hipertensi. Dari hasil *clustering* diperoleh 2 kelompok, yakni jumlah pasien sebanyak 213 orang yang masuk pada cluster pertama sejumlah 69 orang dan cluster kedua sebanyak 144 orang.

Membuat Plot



Gambar 4.2 Plot Cluster

Pada Gambar 4.2, dapat dilihat bahwa daerah yang berwarna biru adalah cluster pertama yang menggambarkan pasien hipertensi dengan tingkat keparahan yang rendah dan daerah yang berwarna hijau adalah cluster kedua yang menggambarkan pasien hipertensi dengan tingkat keparahan yang tinggi.

5. KESIMPULAN

Dari hasil penelitian K-Means Clustering untuk pengelompokan pasien hipertensi dipuskesmas Jongaya yang telah diuraikan disimpulkan bahwa penerapan algoritma K-Means Clustering pada pasien hipertensi dipuskesmas Jongaya, menghasilkan sebuah informasi mengenai data pengelompokan pasien tertinggi dan terendah. Dari 213 pasien yang diteliti terdapat 69 pasien yang merupakan anggota Cluster 1 atau dapat diartikan sebagai pasien yang memiliki tingkat keparahan yang rendah dan 144 pasien yang masuk ke dalam Cluster 2 yang dapat diartikan sebagai pasien yang memiliki tingkat keparahan yang tinggi.

6. DAFTAR PUSTAKA

- [1] Nuraeni, E. (2019). Hubungan Usia Dan Jenis Kelamin Beresiko Dengan Kejadian Hipertensi Di Klinik X Kota Tangerang. *Jurnal JKFT*, 4(1), 1-6.
- [2] Tiara, U. I. (2020). Hubungan obesitas dengan kejadian hipertensi. *Journal of Health Science and Physiotherapy*, 2(2), 167-171.
- [3] Ajeng Sinta Nuryani, A. (2021). Hubungan Merokok Dengan Kejadian Hipertensi.
- [4] Anam, K. (2016). Gaya hidup sehat mencegah penyakit hipertensi. *Jurnal Langsung*, 3(2).
- [5] Saputra, Ari Sigit dkk. 2013. "Pemodelan Mixture Survival". *Jurnal Biometrika dan Kependudukan*, Vol.2, No.1 Juli 2013:76
- [6] Suiraoaka, I. P. (2012). Penyakit degeneratif. Yogyakarta: Nuha Medika, 45(51).
- [7] Darmi, Y. D., & Setiawan, A. (2016). Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama*, 12(2).