

Klasifikasi Penyakit Ginjal dengan Metode K-Means

Andi Sri Irtawaty^{1*}

¹Teknik Elektronika, Politeknik Negeri Balikpapan

*andi.sri@poltekba.ac.id

Abstract

Chronic Kidney Disease is a serious problem in the world. According to WHO, in 2001 and Burden of Disease, was ranked the 12th highest number of deaths (about 850,000 people annual) [1]. Based on these data, so in this research will classify kidney disease into five clusters through the implementation of the K-means method, This method is a method that uses an algorithm that is best in Partitional Clustering algorithms and the most commonly used among other Clustering algorithms, due to its simplicity and efficiency. Test parameters in this study there are three, namely urea, kritinien and GFR. This research will be tested 10 samples of patient data in which the values of the parameters of the test will go through a process of iteration until it reaches a convergent value. The values will be labeled as centroid value for clustering of kidney disease. Clusters consist of normal kidney condition, the symptoms of kidney stage 1, stage 2, stage 3 and stage 4. The accuracy of the test using the K-Means of about 90%.

Keywords: kidney disease, K-means method, urea, kritinien, GFR.

Abstrak

Penyakit Ginjal Kronik merupakan masalah serius di dunia. Menurut WHO, 2001 dan *Burden of Disease*, menduduki peringkat ke-12 tertinggi angka kematian (sekitar 850.000 orang setiap tahunnya) [1]. Berdasarkan data tersebut, maka dalam penelitian ini akan mengklasifikasikan penyakit ginjal menjadi 5 cluster melalui implementasi metode K-means. Metode ini merupakan metode yang menggunakan algoritma yang terbaik dalam algoritma *Partitional Clustering* dan yang paling sering digunakan diantara algoritma pengclusteran lainnya, karena kesederhanaan dan efisiensinya. Parameter uji dalam penelitian ini ada 3 yaitu ureum, kritinien dan GFR. Pada penelitian ini akan diuji 10 sample data pasien dimana nilai-nilai parameter ujinya akan melalui proses iterasi sampai mencapai nilai yang konvergen. Nilai yang telah konvergen akan diberi label sebagai nilai centroid untuk pengclusteran dari penyakit ginjal, Clusternya terdiri atas kondisi ginjal normal, gejala ginjal stadium 1, stadium 2, stadium 3 dan stadium 4. Tingkat akurasi pengclusteran penyakit ginjal menggunakan metode K-Means dalam penelitian ini cukup tinggi, sekitar 90%.

Kata kunci: penyakit ginjal, metode K-means, ureum, kritinien, GFR.

1. Pendahuluan

Di Indonesia, jumlah penderita gagal ginjal kronik terus meningkat dan diperkirakan pertumbuhannya sekitar 10% setiap tahun. Dari data di beberapa pusat nefrologi di Indonesia diperkirakan prevalensi penyakit ginjal kronik masing-masing berkisar 100–150 / 1 juta penduduk. Fungsi ginjal memegang peranan penting dalam tubuh manusia, yaitu sebagai sistem ekskresi (mengeluarkan zat-zat sisa metabolisme yang tidak berguna bagi tubuh). [2]

Ada 3 parameter ukur untuk mengetahui kondisi ginjal yang sehat, yaitu

1. Ureum : kadar ureum pada ginjal yang sehat 15 - 40 mg/dl,
2. Kritinien : kadar kritinien normal yaitu 0,1 – 1,1 mg/dl,
3. Glomerulus Filtration Rate (GFR) atau Laju Filtrasi Glomerulus (LFG): kadar GFR normal yaitu 90 – 120 mL/min/1,73 m². [3]

Gagal ginjal kronik dapat dibagi menjadi 5 stadium berdasarkan nilai Laju Filtrasi Glomerulus (LFG). Glomerulus

adalah struktur di ginjal yang berfungsi melakukan penyaringannya, [4], [5]. Stadiumnya adalah sebagai berikut :

- Stadium 1: GRF_90
- Stadium 2: GRF_(60-89)
- Stadium 3: GRF_(30_59)
- Stadium 4: GRF_(15-29)

Identifikasi dini penyakit gagal ginjal dapat dilakukan dengan metode k-means [6]. K-means adalah salah satu metode cluster analisis non hirarki yang berusaha mempartisi objek yang ada ke dalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya. Pengclusteran objek diperoleh dari jarak objek dengan centroid (titik pusat) yang terdekat [7]. Dalam penelitian ini, akan diuji 10 sample data primer ginjal. Berdasarkan hasil pengolahan data penelitian sebelumnya (Dian, 2016), metode k-means memiliki tingkat akurasi sekitar 82% dalam mengelompokkan data citra [8].

2. Metoda Penelitian

2.1. Metode K-Means

Metode yang digunakan dalam penelitian ini adalah metode K-means. Tahapan dalam penelitian ini meliputi:

1. Tahap I (Persiapan)

Tahap ini meliputi tahapan studi literatur dari permasalahan yang dikemukakan, Studi literatur diperoleh dari buku, jurnal, artikel laporan penelitian dan situs dari internet, Dalam tahapan ini dilakukan proses pengumpulan data primer (data pakar) dari beberapa literatur. [2]

2. Tahap II (Penelitian)

Langkah-langkah dari tahap ini adalah sebagai berikut:

- a) Tentukan jumlah cluster,
- b) Inisialisasi k sebagai pusat cluster (beri nilai-nilai random),
- c) Alokasikan setiap data atau obyek ke cluster terdekat, Kedekatan dua obyek ditentukan berdasarkan jarak antar kedua obyek tersebut, Jarak paling dekat antara satu data dengan satu cluster tertentu

akan menentukan suatu data masuk ke dalam cluster yang mana,

- d) Hitung kembali pusat cluster dengan anggota cluster yang sekarang, Pusat cluster adalah rata-rata semua data atau obyek dalam cluster,
- e) Tugaskan lagi setiap obyek memakai pusat cluster yang baru, Jika pusat cluster sudah tidak berubah lagi, maka proses pengclusteran selesai,
- f) Kembali ke- langkah 3 sampai pusat cluster tidak berubah lagi, (apabila perubahan nilai centroid masih di atas nilai threshold yang ditentukan, atau apabila perubahan pada nilai *objective function* masih di atas nilai threshold yang ditentukan).

Untuk menghitung nilai centroid cluster ke-I, vi, digunakan rumus sebagai berikut:

$$v_{i,j} = \frac{\sum_{k=1}^{N_i} x_{kj}}{N_i} \dots\dots\dots(1)$$

Keterangan :

V_{ij} = nilai centroid

X_{k,j} = koordinat objek

N_i = banyaknya dimensi

Hasil pengolahan data primer disajikan pada tabel 1, 2, 3, 4 dan 5.

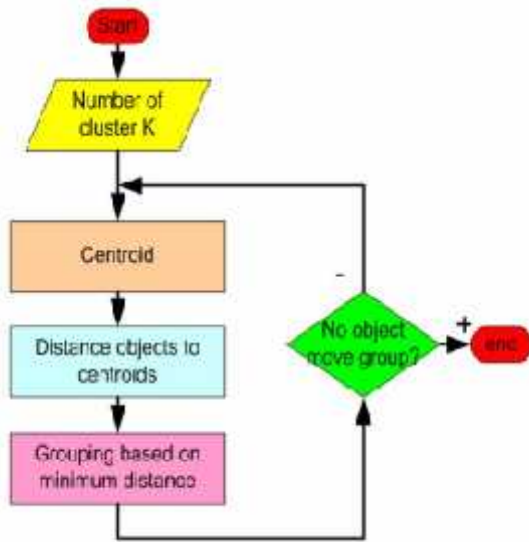
3. Tahap III (Analisis Data)

Pada tahap analisis, hasil pengolahan data primer dari 10 sample, mulai konvergen setelah perhitungan sampai iterasi ke-5.

4. Tahap IV (Kesimpulan)

Berdasarkan hasil analisis (tahap III), diperoleh kesimpulan bahwa tiga orang memiliki ginjal sehat, dua orang terindikasi gejala ginjal stadium 1, tiga orang terindikasi penyakit ginjal stadium 2, satu orang terindikasi penyakit ginjal stadium 3 dan satu orang terindikasi gagal ginjal kronis (stadium 4).

2.2. Perancangan Flowchart K-Means Clustering



Gambar 1. Flowchart K-Means Clustering

Langkah-langkah pengolahan data primer terlihat pada Gambar 1. diatas dengan penjelasan sebagai berikut:

- Berdasarkan study literatur, cluster dibagi menjadi 5 kategori, yaitu:
 - Cluster 1: Normal
 - Cluster 2: Stadium 1
 - Cluster 3: Stadium 2
 - Cluster 4: Stadium 3
 - Cluster 5: Stadium 4
- Inisialisasi k sebagai pusat cluster (nilai parameter ukur ada 3 yaitu ureum, kreatinin dan GFR).
- Alokasikan setiap data atau obyek ke cluster terdekat.

Tabel 1. Menentukan nilai k dan alokasi data ke cluster terdekat

Cluster	Kategori	Ureum	Kreatinin	GFR
1	Normal	27,5	1	105
2	Stadium 1	40	1,5	90
3	Stadium 2	43	2,5	74,5
4	Stadium 3	45	5	44,5
5	Stadium 4	50	9	17,5

- Hitung kembali pusat cluster (*centroid*) dengan anggota cluster yang sekarang.
- Ulangi perhitungan setiap obyek memakai pusat cluster (*centroid*) yang baru, Jika pusat cluster sudah tidak berubah lagi, maka proses pengclustering selesai (pengolahan data konvergen).

Tabel 2. Perhitungan awal centroid

Iterasi 1			Iterasi 2			Iterasi 3		
Ureum	Kreatinin	GFR	Ureum	Kreatinin	GFR	Ureum	Kreatinin	GFR
27,67	1,2	104	27,67	1,2	104	27,67	1,2	104
41,47	2,23	85,7	40,75	1,85	91	40,75	1,85	91
42,83	2,57	80,7	43,75	2,85	75,5	43,75	2,85	75,5
45,33	4,73	56	46	5,75	46	46	5,75	46
47,67	7,17	36,7	51	10	18	51	10	18

- Kembali ke- langkah 3 sampai pusat cluster tidak berubah lagi,

Tabel 3. Perhitungan akhir centroid iterasi 1

Data	Ureum	Kreatinin	GFR	Iterasi 1				
				r1	r2	r3	r4	r5
1	40,5	1,7	91	19,12	1,14	16,71	46,83	74,47
2	28	1,1	106	1,12	12,05	34,92	63,93	91,54
3	29	1,2	102	3,36	16,28	30,89	59,81	87,42
4	43,5	3	75	34,06	15,48	0,87	30,60	58,18
5	46	5,5	45	62,95	45,57	29,80	1,22	28,01
6	41	2	91	19,47	1,50	16,63	46,77	74,38
7	51	10	18	90,57	73,33	57,55	27,63	1,50
8	26	1,3	104	1,83	19,80	34,07	62,57	90,10
9	44	2,7	76	33,41	14,61	1,81	31,60	59,14
10	46	6	47	61,08	43,65	27,88	91,51	29,92

Tabel 4. Perhitungan akhir centroid iterasi 2 dan 3

Iterasi 2					Iterasi 3				
r1	r2	r3	r4	r5	r1	r2	r3	r4	r5
18,27	5,49	10,63	35,46	55,08	18,27	0,29	15,88	45,52	74,22
2,03	5,49	29,39	53,04	100,95	2,03	19,70	34,37	62,81	91,39
2,40	20,70	25,46	48,94	2,40	2,40	16,11	30,37	58,70	87,28
33,09	10,85	5,72	19,17	38,78	33,09	16,28	0,58	29,24	57,92
61,93	41,03	35,93	11,05	8,66	61,93	46,44	30,70	1,03	27,83
18,64	0,29	15,76	45,43	74,11	18,64	0,29	15,76	45,43	74,11
89,54	74,17	58,39	28,76	0	89,54	74,17	58,39	28,76	0
1,67	19,67	33,61	61,51	89,98	18,27	0,29	15,88	45,52	74,22
32,45	15,37	0,58	30,22	58,88	2,03	19,70	34,37	62,81	91,39
60,07	44,51	28,76	1,03	29,70	2,40	16,11	30,37	58,70	87,28

Tabel 5. Perhitungan akhir centroid iterasi 4 dan 5

Iterasi 4					Iterasi 5				
r1	r2	r3	r4	r5	r1	r2	r3	r4	r5
18,27	0,29	15,88	45,52	74,22	18,27	0,29	15,88	45,52	74,22
2,03	19,70	34,37	62,81	91,39	2,03	19,70	34,37	62,81	91,39
2,40	16,11	30,37	58,70	87,28	2,40	16,11	30,37	58,70	87,28
33,09	16,28	0,58	29,24	57,92	33,09	16,28	0,58	29,24	57,92
61,93	46,44	30,70	1,03	27,83	61,93	46,44	30,70	1,03	27,83
18,64	0,29	15,76	45,43	74,11	18,64	0,29	15,76	45,43	74,11
89,54	74,17	58,39	28,76	0	89,54	74,17	58,39	28,76	0
1,67	19,67	33,61	61,51	89,98	1,67	19,67	33,61	61,51	89,98
32,45	15,37	0,58	30,22	58,88	32,45	15,37	0,58	30,22	58,88
60,07	29,70	29,70	1,03	29,70	60,07	29,70	29,70	1,03	29,70

3. Hasil Penelitian

Berdasarkan hasil perhitungan akhir centroid pada Tabel 3, Tabel 4 dan Tabel 5, perhitungan nilai-nilai centroid bernilai konvergen pada iterasi ke-5 pada Tabel 5. Sehingga semua nilai centroid yang terkecil

pada iterasi ke-5 menentukan cluster dari gejala penyakit ginjal. Hasil pengclustering 10 data uji, disajikan pada Tabel 6.

Tabel 6. Hasil Pengclustering

Data	Ureum	Kritinien	GFR	Nilai Centroid	Cluster	Keterangan
1	40,5	1,7	91	0,29	C2	Stadium 1
2	28	1,1	106	2,03	C1	Normal
3	29	1,2	102	2,4	C1	Normal
4	43,5	3	75	0,58	C3	Stadium 2
5	46	5,5	45	1,03	C4	Stadium 3
6	41	2	91	0,29	C2	Stadium 1
7	51	10	18	0	C5	Stadium 4
8	26	1,3	104	1,67	C1	Normal
9	44	2,7	76	0,58	C3	Stadium 2
10	46	6	47	1,03	C3	Stadium 2

Tabel 6. menunjukkan hasil pengclustering untuk 10 sample data uji, yang dikelompokkan sebagai berikut :

- a) Ginjal normal: 3 sample (data 2, data 3 dan data 8).
- b) Ginjal gejala stadium 1: 2 sample (data 1 dan data 6).
- c) Gagal ginjal stadium 2: 3 sample (data 4, data 9 dan data 10).
- d) Gagal ginjal stadium 3: 1 sample (data 5).
- e) Gagal ginjal stadium 4: 1 sample (data 7).

Prosentase akurasi pengclustering dengan metode K-means berdasarkan Tabel 3 dan Tabel 5.

$$\text{Akurasi} = \frac{\text{jumlah data konvergen pada iterasi E dari data yang berbeda pada iterasi}}{\text{jumlah data total yang konvergen pada iterasi E}} \dots (2)$$

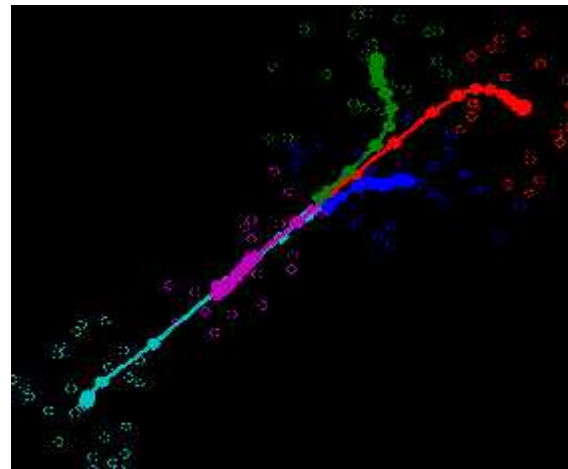
$$= \left(\frac{10-1}{10}\right) \times 100\%$$

$$= 90\%$$

Berdasarkan data pakar pada Tabel 2 perihal nilai normal ureum, kritinien dan GFR, maka tampilan simulasi yang disajikan pada gambar 2, menunjukkan bahwa :

- a) Warna biru: centroid untuk kondisi ginjal normal
- b) Warna ungu: centroid untuk kondisi stadium 1
- c) Warna hijau: centroid untuk kondisi stadium 2

- d) Warna merah: centroid untuk kondisi stadium 3
- e) Warna biru kehijauan: centroid untuk kondisi stadium 4.



Gambar 2. Tampilan simulasi nilai centroid terpendek dengan metode K-means

4. Kesimpulan

Berdasarkan hasil analisa pada penelitian ini, maka dapat ditarik kesimpulan sebagai berikut:

1. Metode K-means terbukti cukup akurat untuk mengclustering penyakit gagal ginjal berdasarkan hasil olahan data primer yang selanjutnya dibandingkan dengan data pakar yang diperoleh dari literatur (tingkat akurasi 90%).
2. Dari 10 sample data diperoleh hasil sebagai berikut:
 - Ginjal normal: 3 sample
 - Ginjal gejala stadium 1: 2 sample
 - Gagal ginjal stadium 2: 3 sample
 - Gagal ginjal stadium 3: 1 sample
 - Gagal ginjal stadium 4: 1 sample

5. Saran

Saran yang diperlukan dalam penelitian ini adalah sebagai berikut:

1. Pada penelitian berikutnya, hasil uji 10 data primer ginjal menggunakan metode K-means akan dikombinasikan dengan metode wavelet dB2.
 2. Jika ditambahkan dengan tampilan citra ginjalnya, maka prosentase pengclusteran penyakit ginjal akan lebih akurat.
 3. Untuk mendapatkan hasil pengujian yang sempurna, maka sebaiknya sample data uji ditambahkan lagi menjadi 30 sample.
- [7] Hervisari, Musdalifah, Sudarsana, "Implementasi Algoritma K-Means Untuk Diagnosa Penyakit Gagal Ginjal Kronis," *Jurnal Ilmiah Matematika dan Terapan*, Vol,11 No,1 Juni 2014, pp.27-35,2014.
- [8] Dian Eka ratnasari, Marji, Lailil Muflikhah, "Pengembangan Metode Klasifikasi Berdasarkan K-Means dan LVQ", *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, Vol,1 Nomor 1, pp.1-4, 2014.

6. Daftar Pustaka

- [1] Putu Filla (2016). CKD Referat [Online]. Available :
http://www.academia.edu/11793185/CKD_REFERAT
- [2] Stefanus Santosa, Agus Widjanarko, Catur Supriyanto, "Model Prediksi Penyakit Ginjal Kronik Menggunakan Radial Basis Function," *Jurnal Pseudocode*, Vol,III Nomor 2,2016.
- [3] Syaiful Azmi. (2011, Nov 1). Stadium penyakit ginjal kronik [Online]. Available :
<http://www.purtierpla.centa.com/stadium-penyakit-ginjal-kronik>.
- [4] dr.Pramita Handayani. (2016, Dec 8). *Reinfokus. Gagal Ginjal Kronik (edisi 1)* [Online]. Available :
<http://www.reindo.co.id/id/knowledge/detail/21/GAGAL-GINJAL-KRONIK>
- [5] Samsilul Azhar, Herlina Latipa Sari, dan Leni Natalia Zulita, "Sistem pakar Penyakit Ginjal Pada Manusia Menggunakan Metode Forward Chaining," *Jurnal Media Infotama*, Vol,10 Nomor 1,2014.
- [6] Abdul Kadir dan Adhi Susanto, "Teori dan Aplikasi Pengolahan Citra", Yogyakarta: ANDI OFFSET, 2011.pp 50-60.