

IMPLEMENTASI ALGORITMA K-MEANS UNTUK DIAGNOSA PENYAKIT GAGAL GINJAL KRONIS

Hervisari¹, S. Musdalifah² dan I W. Sudarsana³

^{1,2,3} Program Studi Matematika Jurusan Matematika FMIPA Universitas Tadulako
Jalan Soekarno-Hatta Km. 09 Tondo, Palu 94118, Indonesia.

¹vhervisari@gmail.com, ²Selvymusdalifah@yahoo.com, ³sudarsanaiwayan@yahoo.co.id

ABSTRACT

Chronic Kidney Disease (CKD) is a medical, social dan economical problems for patients and their family. It becomes complex problem especially in developing countries which have limited source to cost for Chronic Kidney Disease patients. CKD's patient increased every year due to delay in diagnosis, the patient has died of complications prior to diagnosis. Therefore we need an action to anticipate the number of patients with CKD. In this research, K-Means algorithm is implemented to diagnose CKD. The result will be used in diagnosis process. The result are two *cluster*, each *cluster* consisting of chronic condition and the other ones can be recovered.

Keywords : Algorithm K-means, Chronic Kidney Disease (CKD), Cluster, Diagnosis

ABSTRAK

Gagal Ginjal Kronis (GGK) merupakan masalah medik, sosial dan ekonomi bagi pasien dan keluarganya. Hal ini menjadi masalah yang besar khususnya di negara-negara yang sedang berkembang dan memiliki sumber-sumber terbatas untuk membiayai pasien dengan gagal ginjal kronis. Jumlah penderita GGK terus meningkat tiap tahunnya, hal ini disebabkan karena keterlambatan penegakkan diagnosis tersebut. Pasien sudah meninggal akibat komplikasi sebelum adanya penegakkan diagnosis. Oleh karena itu diperlukan suatu tindakan guna mengantisipasi jumlah penderita GGK. Dalam penelitian ini, diimplementasikan algoritma K-Means untuk mendiagnosa penyakit gagal ginjal kronis menggunakan K-Means *cluster*. Hasil hitung dari K-Means *cluster* akan digunakan dalam proses mendiagnosa. Dataset yang digunakan dalam proses uji coba menghasilkan dua buah *cluster*, yang masing-masing *cluster* terdiri dari pasien yang menderita gejala dan penyebab dengan diagnosa parah dan pasien yang masih dapat disembuhkan.

Kata Kunci : Algoritma K-Means, Cluster, Diagnosa, Gagal Ginjal Kronis

I. PENDAHULUAN

Gagal ginjal kronis biasanya akibat akhir dari kehilangan fungsi ginjal lanjut secara bertahap. Gagal ginjal kronis terjadi bila ginjal sudah tidak mampu mempertahankan lingkungan internal yang

konsisten dengan kehidupan dan pemulihan fungsi tidak dimulai. Pada kebanyakan individu transisi dari sehat ke status kronis atau penyakit yang menetap sangat lamban dan menunggu beberapa tahun (Doenges, 1999; 626).

Umumnya GGK disebabkan oleh penyakit ginjal intrinsik difus dan menahun. Glomerulonefritis, hipertensi esensial dan pielonefritis merupakan penyebab paling sering dari gagal ginjal kronik, kira-kira 60%. 47% penderita gagal ginjal kronis yang berusia lebih dari 60 tahun lebih banyak disebabkan karena gangguan metabolik seperti diabetes melitus. Berdasarkan data USRDS tahun 2000, hipertensi merupakan penyebab ESRD yang paling besar, yaitu sebanyak 21% dari total kasus (Sukandar, 2006).

Peningkatan jumlah gagal ginjal disebabkan keterlambatan penegakkan diagnosis penyakit tersebut. Pasien sudah meninggal akibat komplikasi sebelum ada nya penegakan diagnosis (Sudoyo. dkk, 2006). Penyebab keterlambatan penegakkan diagnosis tersebut adalah banyaknya faktor yang berpengaruh terhadap pilihan-pilihan yang ada atau beragamnya variabel. Algoritma K-Means dengan metode *clustering* merupakan salah satu solusi untuk mengatasi masalah tersebut. Algoritma K-Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan centroid, mean atau "means". Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil) (Santosa,2007). Langkah-langkah dalam algoritma K-Means *clustering* adalah :

1. Menentukan jumlah *cluster*.
2. Menentukan nilai centroid.
3. Menghitung jarak antara titik centroid dengan titik tiap objek. Untuk menghitung jarak tersebut dapat menggunakan *euclidean distance*. Adapun rumusnya adalah sebagai berikut

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

Keterangan :

- D adalah *euclidean distance*
- $i = 1, 2, \dots, n$ adalah banyaknya dimensi
- (x, y) adalah koordinat objek

Dari perhitungan jarak diatas, cara penggunaannya adalah memasukkan sebuah objek ke dalam *cluster* tertentu dengan mengukur 'jarak' objek tersebut dengan pusat *cluster*. Jika data ada dalam jarak yang masih dalam batas tertentu (mendekati nilai yang lebih kecil)

maka data tersebut dapat dimasukkan pada *cluster* sedemikian hingga jarak tiap-tiap objek ke pusat kelompok di dalam satu kelompok adalah minimum.

4. Kembali ke tahap 2, lakukan perulangan hingga nilai titik tengah yang di hasilkan tetap dan anggota *cluster* tidak berpindah-pindah ke *cluster* lain.

II. METODE PENELITIAN

Berikut adalah penjelasan dari tahapan penelitian yang akan dilakukan penyusun dalam proses penelitian :

1. Memulai Penelitian.
2. Melakukan studi literatur dengan mengumpulkan materi dari buku-buku, jurnal atau artikel yang di dapat dari perpustakaan dan perpustakaan online.
3. Mengkonfirmasi ke Dokter mengenai gejala penyakit yang akan diteliti.
4. Menentukan diagnosa gejala penyakit gagal ginjal kronis menggunakan metode algoritma K-Means.
5. Mengelompokkan objek berdasarkan jarak minimum.
6. Memperoleh hasil diagnosa.

III. HASIL DAN PEMBAHASAN

3.1. HASIL

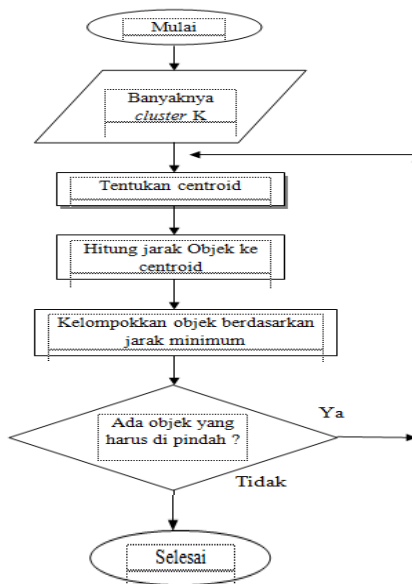
Bab ini menjelaskan tahap-tahap untuk menentukan diagnosa gejala penyakit gagal ginjal kronis menggunakan metode algoritma K-Means. Dengan asumsi yang diberikan jika,

1. Nilai dari suatu gejala ≥ 3 maka diagnosa yang dilakukan untuk penyakit gagal ginjal termasuk parah dan tidak dapat disembuhkan lagi.
2. Nilai dari suatu gejala < 3 maka diagnosa yang dilakukan untuk penyakit gagal ginjal termasuk tidak terlalu parah atau masih dapat di sembuhkan.

3.1.1. Mengkonfirmasi Gejala-Gejala Penyakit

Dari hasil wawancara dengan Spesialis Penyakit Dalam RSUD Undata Palu **dr. Rustam Amirudin Sp. PD** pada tahun 2013, diperoleh 10 gejala yang menjadi faktor penyebab terjadinya resiko. Yaitu, merasa mual berakhir muntah-muntah, berkurangnya kandungan urine saat buang air kecil, merasa lelah dan lemah karena penurunan hemoglobin, hilangnya nafsu makan, menurunnya mental secara signifikan, mempunyai masalah dengan tidur, munculnya rasa gatal, otot kejang dan berkedut, mengalami bengkak pada area kaki, betis dan area yang tidak biasanya dan tekanan darah meningkat karena kelebihan cairan.

3.1.2. Flowchart Algoritma K-Means



3.1.3. Pengambilan Data

Setelah mengkonfirmasi kedokter mengenai gejala dan penyebab terjadinya penyakit gagal ginjal kronis, selanjutnya mengadakan wawancara langsung dengan dokter spesialis penyakit dalam yang menangani penyakit tersebut. Hasil wawancara yang diperoleh adalah sebagai berikut.

Tabel 1 : Data Umum Pasien yang Menderita GGK Tahun 2013

Pasien	Gejala	Penyebab
1.	- Berkurangnya kandungan Urine - Hilangnya nafsu makan - Mengalami bengkak pada area kaki - Tekanan Darah tinggi	- Batu ginjal - Hipertensi
2.	- Tekanan darah tinggi - Mempunyai masalah dengan tidur - Munculnya rasa gatal - Mengalami bengkak pada area kaki - Merasa mual	- Hipertensi - Insomnia - Diabetes
3.	- Berkurangnya kandungan urine - Merasa lelah	- Batu ginjal - Lupus
4.	- Munculnya rasa gatal - Hilangnya nafsu makan - Menurunnya mental - Mengalami gangguan tidur	Lupus

5.	- Merasa mual - Berkurangnya kandungan urine - Hilangnya nafsu makan - Bengkak pada area kaki	- Batu ginjal - Diabetes - Lupus
6.	Berkurangnya kandungan urine	Batu ginjal
7.	- Hilangnya nafsu makan - Mempunyai gangguan tidur	
8.	- Tekanan darah tinggi - Bengkak pada area kaki	- Hipertensi - Diabetes
9.	- Hilangnya nafsu makan - Menurunnya mental - Munculnya rasa gatal	- Lupus - Insomnia
10.	Tekanan darah tinggi	Hipertensi

Dari data di atas, kita memiliki 10 objek sebagai titik data pelatihan dan setiap obyek memiliki 2 atribut mewakili koordinat dari objek, yaitu :

Objek atribut 1 (x) : Gejala

Objek atribut 2 (y) : Penyebab

Tabel 2 : Data Kasus GJK tahun 2013

Objek	Gejala (x)	Penyebab (y)
Pasien A	4	2
Pasien B	5	3
Pasien C	2	2
Pasien D	4	1
Pasien E	4	1
Pasien F	1	3
Pasien G	2	1
Pasien H	2	0
Pasien I	3	2
Pasien J	1	1

3.1.4. Menentukan Pusat *Cluster* dan Jarak Data

1. Menentukan jumlah *cluster*

Dengan memperhatikan data tersebut, kita dapat mengelompokkan objek tersebut ke dalam dua *cluster* sesuai dengan atributnya (yaitu *cluster* 1 dan *cluster* 2). Dari data yang diperoleh, dapat ditentukan bahwa 10 objek tersebut

memiliki 2 atribut (gejala dan penyebab), dimana tiap-tiap pasien mewakili satu titik dengan 2 atribut (x,y) .

2. Menentukan nilai centroid

Untuk menentukan nilai awal centroid dilakukan secara acak. Disini, dimisalkan titik koordinat pasien A, B, C, D, E adalah *cluster* 1 (C_1) dan pasien F, G, H, I, J adalah *cluster* 2 (C_2) sebagai nilai centroid awal.

Tabel 3 : Centroid Awal

Cluster	X_1	X_2
{ A, B, C, D, E}	$\frac{4 + 5 + 2 + 4 + 4}{5} = \frac{19}{5}$	$\frac{2 + 3 + 2 + 1 + 3}{5} = \frac{11}{5}$
{ F, G, H, I, J}	$\frac{1 + 2 + 2 + 3 + 1}{5} = \frac{9}{5}$	$\frac{1 + 0 + 2 + 2 + 1}{5} = \frac{6}{5}$

Sehingga centroid awal yang terbentuk untuk *cluster* berturut-turut adalah $C_1 = (\frac{19}{5}, \frac{11}{5})$ dan $C_2 = (\frac{9}{5}, \frac{6}{5})$.

3. Menghitung jarak antara titik centroid dengan tiap titik objek.

Pada langkah ini, misalkan dengan menggunakan perhitungan kuadrat jarak Euclid, dihitung jarak setiap objek dari pusat *cluster* (centroid) dan memasukkan hasilnya kedalam *cluster* terdekat. Pada setiap pemindahan objek pada suatu *cluster* tertentu, k rataannya harus dihitung kembali sebelum dilanjutkan keproses berikutnya. Perhitungan jaraknya sebagai berikut :

$$d^2 (A, \{A, B, C, D, E\}) = \left(4 - \frac{19}{5}\right)^2 + \left(2 - \frac{11}{5}\right)^2 = (0,2)^2 + (-0,2)^2 = 0,08$$

$$d^2 (A, \{F, G, H, I, J\}) = \left(4 - \frac{9}{5}\right)^2 + \left(2 - \frac{6}{5}\right)^2 = (2,2)^2 + (0,8)^2 = 5,48$$

Karena jarak A lebih dekat dengan *cluster* {A,B,C,D,E} dari pada {F,G,H,I,J}, maka A tidak perlu pindah dari *cluster* {A,B,C,D,E}.

$$d^2 (B, \{A, B, C, D, E\}) = \left(5 - \frac{19}{5}\right)^2 + \left(3 - \frac{11}{5}\right)^2 = (1,2)^2 + (0,8)^2 = 2,08$$

$$d^2 (B, \{F, G, H, I, J\}) = \left(5 - \frac{9}{5}\right)^2 + \left(3 - \frac{6}{5}\right)^2 = (3,2)^2 + (1,8)^2 = 13,48$$

Karena jarak B lebih dekat dengan *cluster* {A,B,C,D,E} dari pada {F,G,H,I,J}, maka B tidak perlu pindah dari *cluster* {A,B,C,D,E}.

$$d^2 (C, \{A, B, C, D, E\}) = \left(2 - \frac{19}{5}\right)^2 + \left(2 - \frac{11}{5}\right)^2 = (-1,8)^2 + (-0,2)^2 = 3,28$$

$$d^2 (C, \{F, G, H, I, J\}) = \left(2 - \frac{9}{5}\right)^2 + \left(2 - \frac{6}{5}\right)^2 = (0,2)^2 + (0,8)^2 = 0,68$$

Karena jarak C lebih dekat dengan *cluster* {F,G,H,I,J} dari pada {A,B,C,D,E}, maka perlu ada pemindahan *cluster* C ke *cluster* {F,G,H,I,J}.

$$d^2(D, \{A, B, C, D, E\}) = \left(4 - \frac{19}{5}\right)^2 + \left(1 - \frac{11}{5}\right)^2 = (0,2)^2 + (-1,2)^2 = 1,48$$

$$d^2(D, \{F, G, H, I, J\}) = \left(4 - \frac{9}{5}\right)^2 + \left(1 - \frac{6}{5}\right)^2 = (2,2)^2 + (-0,2)^2 = 4,88$$

Karena jarak D lebih dekat dengan *cluster* {A,B,C,D,E} dari pada {F,G,H,I,J}, maka D tidak perlu pindah dari *cluster* {A,B,C,D,E}.

$$d^2(E, \{A, B, C, D, E\}) = \left(4 - \frac{19}{5}\right)^2 + \left(3 - \frac{11}{5}\right)^2 = (0,2)^2 + (0,8)^2 = 0,68$$

$$d^2(E, \{F, G, H, I, J\}) = \left(4 - \frac{9}{5}\right)^2 + \left(3 - \frac{6}{5}\right)^2 = (2,2)^2 + (1,8)^2 = 8,08$$

Karena jarak E lebih dekat dengan *cluster* {A,B,C,D,E} dari pada {F,G,H,I,J}, maka E tidak perlu pindah dari *cluster* {A,B,C,D,E}.

Dengan cara yang sama diperoleh pengelompokkan sebagai berikut:

F → jarak F lebih dekat ke *cluster* {F,G,H,I,J} dari pada ke {A,B,C,D,E}, maka tidak terjadi perpindahan *cluster*.

G → Karena jarak G lebih dekat ke *cluster* {F,G,H,I,J} dari pada ke {A,B,C,D,E}, maka tidak terjadi perpindahan *cluster*.

H → Karena jarak H lebih dekat ke *cluster* {F,G,H,I,J} dari pada ke {A,B,C,D,E}, maka tidak terjadi perpindahan *cluster*.

I → Karena jarak I lebih dekat ke *cluster* {A,B,C,D,E} dari pada ke {F,G,H,I,J}, maka perlu ada perpindahan *cluster*.

J → Karena jarak J lebih dekat ke *cluster* {F,G,H,I,J} dari pada ke {A,B,C,D,E}, maka tidak terjadi perpindahan *cluster*.

Dari perhitungan diatas, dapat dilihat jarak yang terbentuk dari centroid awal yaitu *cluster* 1 {A,B,D,E,I} dan *cluster* yang ke 2 {C,F,G,H,J}. Sehingga masih terdapat *cluster* yang tidak menempati anggotanya, maka perlu untuk membentuk centroid baru.

Tabel 4 : Centroid Baru

<i>Cluster</i>	X_1	X_2
{A, B, D, E, I}	$\frac{4 + 5 + 4 + 4 + 3}{5} = \frac{20}{5} = 4$	$\frac{2 + 3 + 1 + 3 + 2}{5} = \frac{11}{5}$
{C, F, G, H, J}	$\frac{2 + 1 + 2 + 2 + 1}{5} = \frac{8}{5}$	$\frac{2 + 1 + 0 + 2 + 1}{5} = \frac{6}{5}$

Sehingga centroid baru yang terbentuk adalah $C_1 = \left(4, \frac{8}{5}\right)$ dan $C_2 = \left(\frac{11}{5}, \frac{6}{5}\right)$. Selanjutnya menghitung kembali jarak antara setiap objek dengan centroid baru, dan masukkan hasilnya ke dalam *cluster* yang terdekat. Perhitungannya sebagai berikut

$$d^2 (A, \{A, B, D, E, I\}) = (4 - 4)^2 + (2 - \frac{11}{5})^2 = (0)^2 + (-0,2)^2 = 0,04$$

$$d^2 (A, \{C, F, G, H, J\}) = (4 - \frac{8}{5})^2 + (2 - \frac{6}{5})^2 = (2,4)^2 + (0,8)^2 = 6,4$$

Karena jarak A lebih dekat dengan *cluster* {A,B,D,E,I} dari pada {C,F,G,H,J}, maka A tidak perlu pindah dari *cluster* {A,B,D,E,I}.

$$d^2 (B, \{A, B, D, E, I\}) = (5 - 4)^2 + (3 - \frac{11}{5})^2 = (1)^2 + (0,8)^2 = 1,64$$

$$d^2 (B, \{C, F, G, H, J\}) = (5 - \frac{8}{5})^2 + (3 - \frac{6}{5})^2 = (3,4)^2 + (1,8)^2 = 14,8$$

Karena jarak B lebih dekat dengan *cluster* {A,B,D,E,I} dari pada {C,F,G,H,J}, maka B tidak perlu pindah dari *cluster* {A,B,D,E,I}.

$$d^2 (C, \{A, B, D, E, I\}) = (2 - 4)^2 + (2 - \frac{11}{5})^2 = (-2)^2 + (-0,2)^2 = 4,04$$

$$d^2 (C, \{C, F, G, H, J\}) = (2 - \frac{8}{5})^2 + (2 - \frac{6}{5})^2 = (0,4)^2 + (0,8)^2 = 0,8$$

Karena jarak C lebih dekat dengan *cluster* {C,F,G,H,J} dari pada {A,B,D,E,I}, maka C tidak perlu pindah dari *cluster* {C,F,G,H,J}.

$$d^2 (D, \{A, B, D, E, I\}) = (4 - 4)^2 + (1 - \frac{11}{5})^2 = (0)^2 + (-1,2)^2 = 1,44$$

$$d^2 (D, \{C, F, G, H, J\}) = (4 - \frac{8}{5})^2 + (1 - \frac{6}{5})^2 = (2,4)^2 + (-0,8)^2 = 5,8$$

Karena jarak D lebih dekat dengan *cluster* {A,B,D,E,I} dari pada {C,F,G,H,J}, maka D tidak perlu pindah dari *cluster* {A,B,D,E,I}.

$$d^2 (E, \{A, B, D, E, I\}) = (4 - 4)^2 + (3 - \frac{11}{5})^2 = (0)^2 + (0,8)^2 = 0,64$$

$$d^2 (E, \{C, F, G, H, J\}) = (4 - \frac{8}{5})^2 + (3 - \frac{6}{5})^2 = (2,4)^2 + (1,8)^2 = 9$$

Karena jarak E lebih dekat dengan *cluster* {A,B,D,E,I} dari pada {C,F,G,H,J}, maka E tidak perlu pindah dari *cluster* {A,B,D,E,I}.

F → Karena jarak F lebih dekat dengan *cluster* {C,F,G,H,J} dari pada {A,B,D,E,I}, maka F tidak perlu pindah dari *cluster* {C,F,G,H,J}.

G → Karena jarak G lebih dekat dengan *cluster* {C,F,G,H,J} dari pada {A,B,D,E,I}, maka G tidak perlu pindah dari *cluster* {C,F,G,H,J}.

H → Karena jarak H lebih dekat dengan *cluster* {C,F,G,H,J} dari pada {A,B,D,E,I}, maka H tidak perlu pindah dari *cluster* {C,F,G,H,J}.

J → Karena jarak J lebih dekat dengan *cluster* {C,F,G,H,J} dari pada {A,B,D,E,I}, maka J tidak perlu pindah dari *cluster* {C,F,G,H,J}.

Terlihat dari proses diatas, tidak terjadi adanya perpindahan *cluster* dan setiap objek sudah menunjuk ke suatu kelompok dengan rataan terdekat, sehingga proses berhenti sampai disitu. Dan *cluster* akhir yang terbentuk adalah

Cluster 1 { A, B, D, E, I }

Cluster 2 { C, F, G, H, J }

3.2. Pembahasan

Dari data hasil perhitungan diatas diperoleh untuk gejala penyakit yang terdapat jumlah paling besar atau ≥ 3 adalah gejala penyakit dalam cluster 1 : { A, B, D, E, I } (Gejala penyakit dapat dilihat pada tabel 1), seperti yang telah di asumsikan bahwa diagnosa penyakit dalam *cluster* 1 adalah gejala penyakit yang parah dan tidak dapat di sembuhkan lagi ginjal yang di miliki. Lalu kemudian pada *cluster* 2 : { C, F, G, H, J } (Gejala penyakit dapat dilihat pada tabel 1) atau pasien yang memiliki gejala penyakit gagal ginjal < 3, termasuk ke dalam diagnosa penyakit gagal ginjal yang masih dalam tahap yang masih dapat disembuhkan.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa Algoritma K-Means dapat digunakan untuk mendiagnosa penyakit gagal ginjal kronis, dimana hasil penelitian menunjukkan bahwa dari 10 pasien, ada 5 pasien yang menderita penyakit gagal ginjal kronis dan 5 pasien lagi dengan diagnosa gagal ginjal yang masih dapat disembuhkan.

DAFTAR PUSTAKA

- [1]. Doenges E, Marilyn, dkk. 1999. *Rencana Asuhan Keperawatan : Pedoman Untuk Perencanaan dan Pendokumentasian Perawatan Pasien*. Edisi 3. EGC. Jakarta.
- [2]. Santoso, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta.
- [3]. Sudoyo, Aru W, dkk. 2006. *Ilmu Penyakit Dalam*. Jilid 3. Edisi 4. Pusat Penerbitan Departemen Ilmu Penyakit Dalam Fakultas Kedokteran Universitas Indonesia. Jakarta.
- [4]. Sukandar, Enday. 2006. *Gagal Ginjal dan Panduan Dialisis*. PPI FK UNPAD. Bandung.