**Analisa Pola Diagnosa Penyakit Ginjal Kronik Algoritma**

***K-means***

**Dea An Nisa 1), Warnia Nengsih, S.Kom., M.Kom.2)**

1. Jurusan Komputer,Politeknik Caltex Riau,Pekanbaru 28265,email: dea14si@mahasiswa.pcr.ac.id
2. Jurusan Komputer, Politeknik Caltex Riau, Pekanbaru 28265, email: warnia.pcr.ac.id

**Abstrak –** Gagal ginjal adalah kondisi saat fungsi ginjal mulai menurun secara bertahap. Indonesia Renal Registry mendefinisikan gagal ginjal kronis sebagai kerusakan ginjal, dapat berupa kelainan jaringan, komposisi darah dan urine atau tes pencitraan ginjal, yang dialami lebih dari tiga bulan. Beberapa kondisi seperti diabetes dan [tekanan darah tinggi](http://www.alodokter.com/hipertensi) menjadi beberapa [penyebab terjadinya gagal ginjal kronis.](http://www.alodokter.com/gagal-ginjal-kronis/penyebab" \t "_blank) Dalam jangka panjang, kondisi-kondisi ini menyebabkan kerusakan pada ginjal sehingga fungsi ginjal menurun. Jika kondisi ini diabaikan, besar kemungkinan terjadi komplikasi pada pasien yang mengalami penyakit ini. Oleh karena itu perlu adanya tindak lanjut yang harus dilakukan untuk mengurangi penderita penyakit gagal ginjal kronik. Pada penelitian ini, diterapkan algoritma *k-means* yang akan membentuk sebuah pola yang hasilnya akan digunakan untuk melakukan prediksi pada pasien untuk mengetahui kondisi pasien yang menderita penyakit gagal ginjal kronik dan pasien yang tidak menderita penyakit gagal ginjal kronik. Dalam proses uji coba, akan digunakan dua buah *cluster*. Analisa ini dilakukan dengan 2 cara, yaitu perhitungan secara manual dan perhitungan dengan bantuan *tools* MatLab. Dengan adanya penelitian ini, masyarakat dapat melihat kemungkinan-kemungkinan untuk terkena penyakit gagal ginjal kronik sehingga dapat diatasi dengan lebih cepat.

**KataKunci** : *K-means*,Gagal Ginjal Kronik, *Matrix Laboratory(MatLab)*

***Abstract -*** Chronic kidney disease is a condition when kidney function begins to decrease gradually. Indonesia Renal Registry defines chronic kidney disease as kidney damage, may be tissue disorders, blood and urine composition or renal imaging tests, which lasts for more than three months. Some conditions such as diabetes and high blood pressure are some of the causes of chronic renal failure. In the long run, these conditions cause damage to the kidneys so that kidney function decreases. If this condition is neglected, there is a high probability of complications in patients with this disease. Therefore, the need for follow-up should be done to reduce chronic kidney failure patients. In this study, k-means algorithm is applied which will form a pattern that the results will be used to predict the patient to know the condition of patients suffering from chronic renal failure and patients who do not suffer from chronic renal failure. In the trial process, two clusters will be used. This analysis is done in 2 ways, which is manual calculation and calculation with the help of MatLab tools. With this research, people can see possibilities for chronic kidney failure so that it can be treated more quickly.

***Keywords:*** *K-means*, *Chronic Kidney Disease*, *Matrix Laboratory(MatLab)*

# **PENDAHULUAN**

Gagal ginjal merupakan salah satu penyakit yang terjadi sebab kelalaian pasien menjaga kesehatan ginjal dan pola hidup yang tidak sehat. Menurut Kementrian Kesehatan Republik Indonesia (Kemenkes) , beberapa tahun belakangan ini tren penyakit ginjal, jantung, penyakit syaraf, kanker, diabetes mellitus, dan haemophilia meningkat setiap tahunnya.

Penyakit ginjal sebenarnya bukanlah penyakit yang dapat langsung di deteksi keberadaannya karena penyakit ini bersifat Nefropati (penyakit tidak menular yang dapat dicegah). Dengan memakan makanan yang tidak melebihi batas kebutuhan nutrisi tubuh dan selalu rutin melakukan olahraga penyakit ini dapat dihindari.

Umumnya penyakit ginjal kronik disebabkan oleh penyakit ginjal intrinsik difus dan berlangusng selama bertahun-tahun. Glomerulonefritis, hipertensi esensial dan pielonefritis merupakan penyebab paling sering dari penyakit ginjal kronik, kira-kira 60%. 47% penderita penyakit ginjal kronis yang berusia di atas 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). Mayoritas penyakit ginjal ini semakin memburuk disebabkan oleh lambatnya kesadaran masyarakat akan pentingnya penanganan penyakit ginjal sejak dini. Sungguh sangat di sayangkan apabila ginjal sudah terlanjur tidak berfungsi menyaring darah dengan semestinya lagi. Biaya yang harus dikeluarkan tidaklah sedikit.

Untuk membantu pendeteksian penyakit ginjal, metode K-Means dengan teknik *cluster* akan sangat membantu. Untuk melakukan metode K-Means *clustering*, data yang dibutuhkan adalah berupa variabel-variabel penyebab (dalam kasus ini penyebab penyakit ginjal). Diharapkan hasil perhitungan K-Means ini dapat diketahui kelompok penyakit ginjal berdasarkan gejala yang dirasakan pasien.

# **LANDASAN TEORI**

## **Data Mining**

Data mining merupakan proses penggalian informasi yang tersembunyi dalam suatu kumpulan data yang besar untuk penemuan pola[1]. Data mining dapat mengotomatisasi prediksi trend, sifat bisnis dan penemuan pola. Namun, data mining mempunyai kendala dalam database dan data mining tidak dapat melakukan analisa sendiri .

*Data Mining* memiliki beberapa tahapan proses data mining yaitu :

* + - 1. Menentukan terlebih dahulu masalah bisnis berdasarkan masalah yang sudah ada
      2. Menentukan teknik, metode dan algoritma yang tepat
      3. Persiapan data
      4. Melakukan proses data.
      5. Validasi berupa evaluasi sesuai atau tidaknya metode yang digunakan.

*Data Mining* memiliki beberapa tahap untuk menemukan pengetahuan dan pola, yaitu :

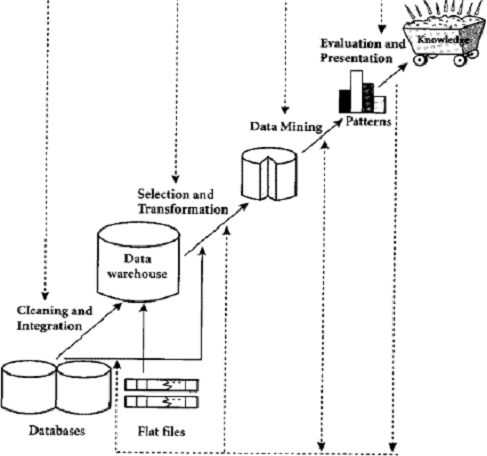
*Data Selection,* pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada sampel data dimana penemuan akan dilakukan.

*Preprocessing/Cleaning,* pada tahap dilakukan pembersihan data noise, duplikasi data, memperbaiki kesalahan data. Output dari *preprocessing* ini adalah data yang sudah bersih dan siap digunakan untuk proses selanjutnya.

*Transformation,* padatahap ini dilakukan penyesuaian pemodelan data agar sesuai dengan format. Konversi tipe data string. Proses ini mentranformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses *mining* dengan cara melakukan agregasi.

*Data Mining,* pada tahap ini data akan diolah dengan metode yang sesuai. Proses ini bisa mencari pola atau informasi dalam data yang sudah dipilih.

*Evaluasi,* pada proses ini akan dilakukan pengujian untuk mengukur akurasi berdasarkan metode yang diambil.

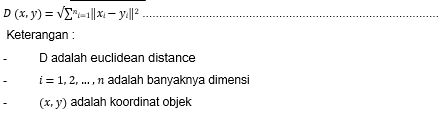


Gambar 2‑1 Tahap Data Mining

## **K-Means**

Algortima 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 :



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.

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

# **METODE PENILITIAN**

## **Preprocessing Data**

*Preprocessing data* merupakan proses untuk melakukan pembersihan pada data. Data mentah biasanya memiliki *null value, redudancy data,*ataupun data yang bersifat *dummy data.* Dengan adanya tahap *preprocessing* ini data yang termasuk *dirty data* akan dibersihkan. Setelah dibersihkan, *outputnya* adalah data yang bersih yang dapat digunakan untuk proses selanjutnya.

Dalam penelitian ini, data yang didapat sudah melewati tahap preprocessing sehingga data sudah siap untuk digunakan karena tidak ada lagi *null value* ataupun *redudancy data.*

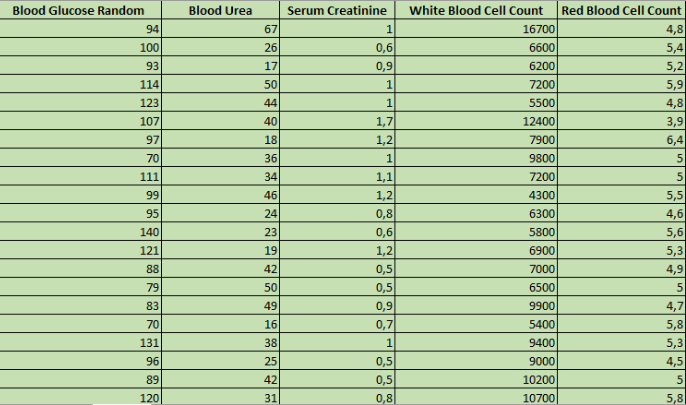
## **Implementasi *K-means***

Data dalam penelitian ini sudah siap untuk digunakan, sehingga langkah selanjutnya yang harus dilakukan adalah tahap perhitungan manual dan perhitungan yang diimplementasikan dengan bantuan matlab.

# **HASIL DAN PEMBAHASAN**

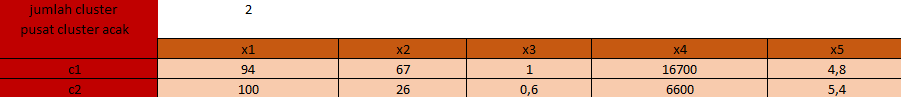
## **Hasil dengan Perhitungan Manual**

Untuk menemukan pola yang akan digunakan dalam mendiagnosa penyakit ginjal kronik, dapat dilakukan dengan cara perhitungan manual. Hal pertama yang harus disiapkan adalah data-data atau atribut yang menjadi penyebab terjadinya penyakit ginjal kronik. Berikut adalah cuplikan data yang akan digunakan :



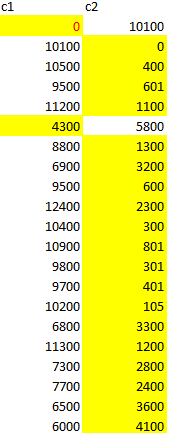
Gambar 4‑1 Tabel Data

Setelah itu tahap selanjutnya adalah menentukan *cluster* acak dan menghitung jarak pusat *cluster.* Pusat *cluster* acak ditentukan secara acak tanpa menggunakan rumus. Angka yang terdapat dalam *cluster* acak harus merupakan anggota dari data atribut(data X) . Dalam satu baris *cluster* boleh memiliki angka yang sama, namun tidak pada satu kolom. Berikut adalah *cluster* acak dari data penelitian penyakit gagal ginjal kronik :



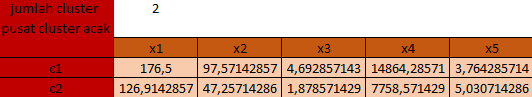
Gambar 4‑2 Cluster Acak

Setelah itu lakukan perhitungan pada dua buah *cluster* tersebutdengan menjumlahkan masing-masing data X. Berikut adalah cuplikan data perhitungan untuk iterasi ke-1 :



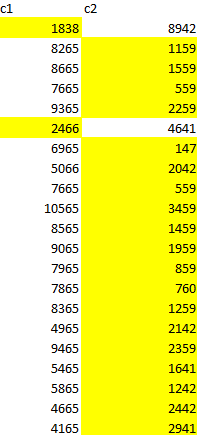
Gambar 4‑3 Iterasi ke-1

Kemudian lakukan iterasi ke-2 dengan mengganti pusat *cluster*. Pada iterasi ke-2, pusat *cluster* tidak ditentukan secara acak, tetapi menggunakan rumus. Berikut adalah cuplikan datanya :



Gambar 4‑4 Pusat Cluster ke-2

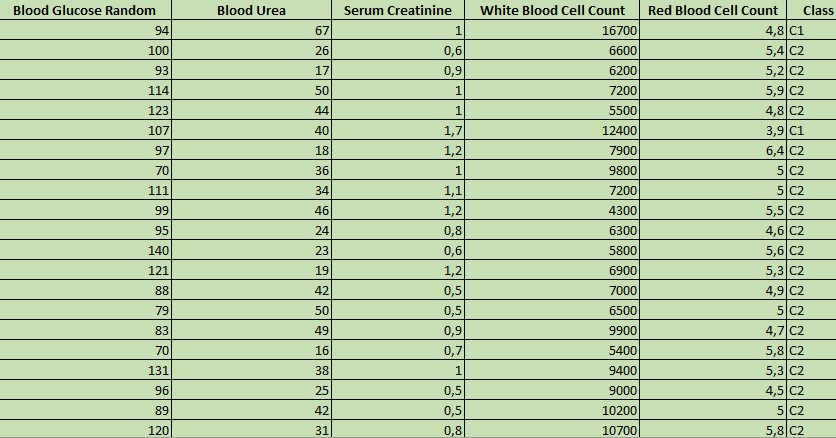
Kemudian lakukan lagi perhitungan dua buah *cluster* pada masing-masing nilai X. Berikut adalah hasil iterasi ke-2 :



Gambar 4‑5 Iterasi ke-2

Jika dibandingkan dengan hasil iterasi ke-1, maka iterasi ke-1 dan iterasi ke-2 sudah memiliki cluster yang sama, sehingga iterasi bisa dihentikan.

Berdasarkan hasil perhitungan di atas, dapat disimpulkan bahwa :

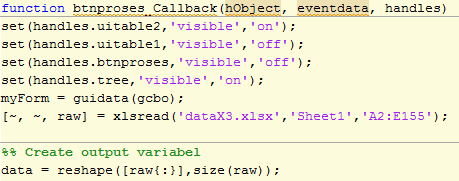


Gambar 4‑6 Tabel dengan Nilai Cluster

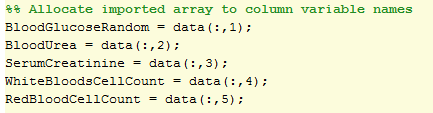
## **Hasil dengan Matlab**

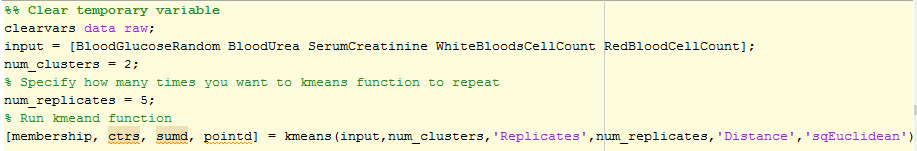
Dengan menggunakan bantuan MatLab, penemuan pola *cluster* akan lebih mudah. Karena tidak perlu melakukan perhitungan panjang untuk mendapatkan hasil *cluster.* Berikut adalah langkah-langkah yang dilakukan dalam penelitian ini untuk menemukan *cluster* pada Matlab :

Buat sebuah button dan tabel yang akan digunakan untuk menampung data yang akan di-*cluster*. Kemudian pada button *callback*, ketikkan syntax berikut :

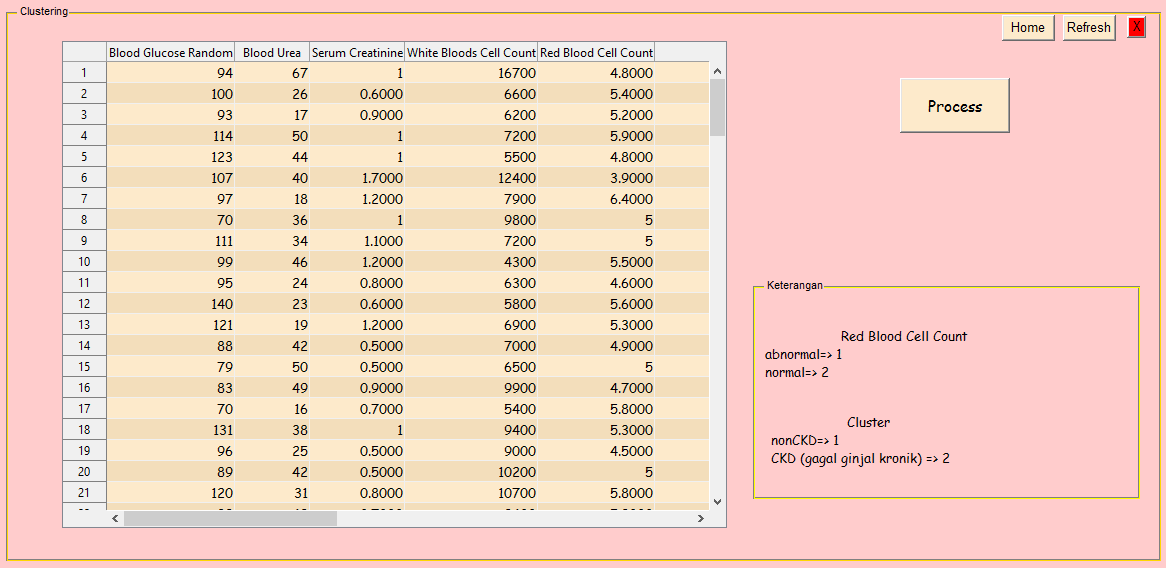


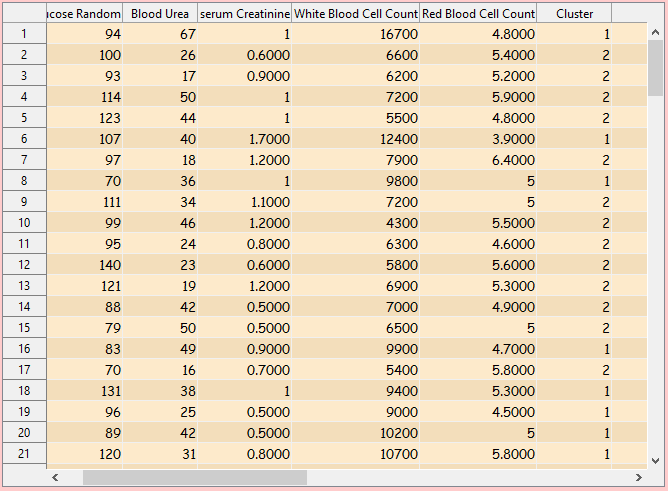
**(‘dataX3.xlsx’,’Sheet1’,’A2:E155’)** merupakan data excel yang akan di­-*cluster.*

Syntax di atas adalah pendeklarasian variabel untuk mengisi kolom-kolom tabel yang sudah dibuat.



**Input**berisi nama-nama variabel yang telah dideklarasikan pada tahap sebelumnya. ­**num\_clusters** menunjukkan berapa cluster yang akan dikerjakan. Pada penelitian ini, hanya akan ada 2 cluster.**Num\_replicates** merupakan berapa kali batas pengulangan cluster. **Kmeans** merupakan syntax untuk menemukan cluster pada dataX. Berikut adalah hasil dari syntax yang telah dibuat :

****



# **KESIMPULAN**

Berdasarkan hasil analisis dan pengerjaan dengan cara perhitungan manual dan pengimplementasian di dalam matlab, penulis mendapatkan hasil cluster yang berbeda. Menurut penulis, analisa yang dilakukan oleh matlab lebih mudah dan kemungkinan terjadi kesalahan lebih sedikit. Sedangkan pada perhitungan manual, resiko terjadi kesalahan dalam perhitungan lebih besar. Hal ini dikarenakan lebih besarnya terjadi *human errror* dan satu kesalahan dalam menghitung angka akan menyebabkan kesalahan sampai akhir perhitungan penentuan cluster. Dari hasil cluster tersebut akan ditemukan pola, jika termasuk cluster1 berarti pasien mengalami penyakit gagal ginjal kronik, sedangkan cluster2 menunjukkan pasien yang tidak mengalami penyakit gagal ginjal kronik

.

# **DAFTAR REFERENSI**

[1]. Sukandar, Enday. 2006. Gagal Ginjal dan Panduan Dialisis.PPI FK UNPAD. Bandung.