

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Definisi Sistem**

Definisi sistem berkembang sesuai konteks dimana pengertian sistem itu digunakan. Berikut akan diberikan definisi sistem secara umum:

1. Kumpulan dari bagian-bagian yang bekerja sama untuk mencapai tujuan yang sama (Jogiyanto, H.M. 2005).
2. Sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu (Jogiyanto, H.M. 2005).
3. Suatu sistem adalah jaringan kerja dari prosedur-prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan suatu kegiatan atau untuk menyelesaikan suatu sasaran yang tertentu (Jogiyanto, H.M. 2005).
4. Sistem merupakan kumpulan elemen-elemen yang saling terkait dan bekerja sama untuk memproses masukan (*input*) yang ditujukan kepada sistem tersebut dan mengolah masukan tersebut sampai menghasilkan keluaran (*output*) yang diinginkan (Kadir A, Triwahyuni TCH. 2003).

Suatu sistem memiliki karakteristik atau sifat-sifat tertentu, yaitu:

1. Komponen sistem (*components*)
2. Batas sistem (*boundary*)
3. Lingkungan luar sistem (*environments*)
4. Penghubung sistem (*interface*)
5. Masukan sistem (*input*)
6. Keluaran sistem (*output*)
7. Pengolah sistem
8. Sasaran sistem (*objective*)

## 2.2 Gagal Ginjal

Ginjal merupakan organ penting yang berfungsi menjaga komposisi darah dengan mencegah menumpuknya limbah dan mengendalikan keseimbangan cairan dalam tubuh, menjaga level elektrolit seperti sodium, potasium dan fosfat tetap stabil, serta memproduksi hormon dan enzim yang membantu dalam mengendalikan tekanan darah, membuat sel darah merah dan menjaga tulang tetap kuat (Kemenkes,2017)

Gagal Ginjal adalah jenis penyakit dimana ginjal kehilangan kemampuannya untuk mempertaruhkan volume dan komposisi cairan tubuh dalam keadaan asupan makanan normal (Price & Wilson, 2006). Penyakit gagal ginjal disebabkan oleh fungsi organ ginjal yang mengalami penurunan, sehingga tidak dapat menyaring pembuangan elektronik tubuh. Selain itu organ ini juga tidak dapat menjaga keseimbangan antara cairan dan zat kimia tubuh, seperti sodium dan kalium di dalam darah atau produksi urine (As'adi M, 2012).

Gagal ginjal kronis adalah proses kerusakan ginjal selama rentan waktu lebih dari tiga bulan, Gagal Ginjal kronis dapat menimbulkan *sintoma* yaitu laju filtrasi glomerular di bawah  $60 \text{ ml}/\text{men}/1.73 \text{ m}^2$  atau di atas nilai tersebut yang disertai dengan kelainan sedimen urine, selain itu adanya batu ginjal juga dapat menjadi indikasi gagal ginjal kronis pada penderita bawaan (As'adi M, 2012). Kegagalan gagal ginjal kronis terjadi bila ginjal sudah tidak mampu mempertahankan lingkungan internal yang konsisten dengan kehidupan dan pemulihan fungsi yang tidak dimulai. Pada banyak kasus transisi dari kondisi sehat ke status kronis (penyakit yang menetap) sangat lamban, bahkan membutuhkan waktu selama beberap tahun (Barbara C. Long, 1996)

## 2.3 Jenis-Jenis Gagal Ginjal

Gagal Ginjal biasanya dibagi menjadi dua kategori yaitu kronik dan akut. Gagal ginjal kronik merupakan perkembangan gagal ginjal yang progresif dan lambat pada setiap *nefron* atau biasanya berlangsung

beberapa tahun dan tidak *reversible*. Gagal ginjal akut seringkali berkaitan dengan penyakit kritis, berkembang cepat dalam hitungan beberapa hari hingga minggu, dan biasanya *reversible* bila pasien dapat bertahan dengan penyakitnya (Price & Wilson, 2006)

#### 1. Gagal Ginjal Kronis

Menurut (Amin H.N & Hardi K, 2015) perjalanan klinisnya penyakit gagal ginjal kronik ada beberapa tahap diantaranya.

- a. Menurunnya cadangan ginjal pasien asimtomatik, namun *GFR* dapat menurun hingga 25% dari normal.
- b. *Infusiasi* ginjal selama keadaan ini pasien mengalami *poliuria* dan *nokturia*, *GFR* 10% hingga 25% dari normal, kadar *serum creatinin* dan *BUN* sedikit meningkat diatas normal.
- c. Penyakit ginjal kronik stadium akhir atau sindrom uremik akan mengalami kejang-kejang sampai koma, yang di tandai dengan *GFR* kurang dari 5 sampai 10 ml/menit, kadar *serum creatinin* dan *BUN* meningkat tajam.

#### 2. Gagal Ginjal Akut

Menurut (Amin H.N & Hardi K, 2015) perjalanan klinis gagal ginjal akut biasanya di bagi menjadi tiga stadium yaitu *oliguria*, *dieresis* dan pemulihan. Tetapi perlu di ingat bahwa gagal ginjal akut dapat saja terjadi saat keluaran urine lebih dari 400ml/24 jam

- a. Stadium *oliguria* timbul dalam waktu 24 – 48 jam sesudah trauma dan disertai *azotemia*.
- b. Stadium *dieresis* timbul gejala keluaran urine lebih dari 400ml/hari berlangsung dua sampai tiga minggu, pengeluaran urin harian, tingginya kadar *urea* darah, kemungkinan pasien yang mengalami akan kekurangan kalium, natrium dan air pada tubuh, selama Stadium *dieresis* kadar *BUN* mungkin akan meningkat terus.

- c. Stadium penyembuhan gagal ginjal akut berlangsung selama satu tahun dan selama itu anemia dan kemampuan pemekatan ginjal sedikit demi sedikit membaik.

#### 2.4 Klasifikasi Penyakit Gagal Ginjal Kronis

Klasifikasi gagal ginjal kronik didasarkan atas dua hal yaitu, atas dasar derajat (stadium) penyakit dan atas dasar diagnosis etiologi (Sudoyo et al, 2009). Klasifikasi atas dasar penyakit, dibuat atas dasar LFG.

**Tabel 2.1** Klasifikasi gagal ginjal kronis berdasarkan stadium

Stadium	Penjelasan	LFG ml/men/1.73 m <sup>2</sup>
1	Kerusakan ginjal dengan LFG normal	≥ 90
2	Kerusakan ginjal dengan LFG menurun ringan	60 - 89
3	Kerusakan ginjal dengan LFG menurun sedang	30 - 59
4	Kerusakan ginjal dengan LFG menurun berat	15 - 29
5	Gagal ginjal kronis	< 15

**Sumber :** (*World Health Organization*).

**Tabel 2.2** Klasifikasi gagal ginjal kronis atas dasar diagnosa etiologi

Penyakit	Tipe Mayor
Penyakit Diabetes	Diabetes tipe 1 dan 2
Penyakit Diabetes	Penyakit glomerular (infeksi sistemik, obat, neoplasia) Penyakit vascular (penyakit pembuluh darah besar, hipertensi, mikroangiopati) Penyakit tubulointerstitial (pielonefritis kronik, batu, obstruksi) Penyakit kistik (ginjal polikistik)
Penyakit transplantasi	Rejeksi kronik Keracunan Obat

**Sumber :** (*Sudoyo et al, 2009*).

## 2.5 Faktor Risiko Penyakit Gagal Ginjal Kronis

Faktor resiko pasien gagal ginjal kronis digolongkan menjadi dua, yaitu faktor resiko yang tidak dapat dirubah seperti faktor usia, jenis kelamin, Pendidikan dan genetik, dan faktor resiko yang dapat dirubah seperti *hemoglobin*, kreatinin atau *serum creatinin*, berat badan, dan *blood urea nitrogen* (BUN).

Faktor resiko yang tidak dapat dirubah meliputi :

### 1. Usia

Gagal ginjal kronis sering menyerang pada orang yang telah lanjut usia, begitu pula ginjal orang yang secara jangka waktu lama akan mengalami penurunan jumlah *nefron* dan LFG atau *GFR* . Hal ini meningkatkan risiko terjadinya gagal ginjal kronis pada lansia. Sebagian besar usia pasien GGK di RS Muhammadiyah Gresik adalah usia lebih dari 40 tahun..

### 2. Jenis Kelamin

Proporsi jenis kelamin sangat berpengaruh pada penyakit gagal ginjal kronis. Dimana data yang di dapat dari RS Muhammadiyah Gresik adalah pasien yang berjenis kelamin laki-laki usia lebih banyak dibanding dengan pasien yang berjenis kelamin wanita.

Sedangkan faktor resiko yang dapat diubah meliputi:

### 1. *Hemoglobin*

Hemoglobin adalah komponen yang berfungsi sebagai alat transportasi oksigen (O<sub>2</sub>) dan karbon dioksida (CO<sub>2</sub>). Hb tersusun dari globin (empat rantai protein yang terdiri dari dua unit alfa dan dua unit beta) dan hemo (mengandung atom besi dan porphyrin: suatu pigmen merah). Satu gram hemoglobin mengangkut 1,34 mL oksigen. Kapasitas angkut ini berhubungan dengan kadar Hb bukan jumlah sel darah merah (Pedoman data interpretasi data klinik, 2011).

**Tabel 2.3** Nilai normal *Hemoglobin*

Nilai Normal	Penjelasan
Laki – Laki	13 - 18 g/dL
Wanita	12 - 16 g/dL

2. *Blood Urea Nitrogen* (BUN)

Pengukuran *Blood Urea Nitrogen* (BUN) atau ureum serum dapat dipergunakan untuk mengevaluasi fungsi ginjal, status hidrasi, menilai keseimbangan nitrogen, menilai progresivitas penyakit ginjal, dan menilai hasil hemodialisis.

**Tabel 2.4** Nilai normal *Blood Urea Nitrogen* (BUN)

Nilai Normal	6 - 20 mg/dL
--------------	--------------

3. Kreatinin (*creatinine*)

Kreatinin dihasilkan selama kontraksi otot skeletal melalui pemecahan kreatinin fosfat. Kreatinin diekskresi oleh ginjal dan konsentrasinya dalam darah sebagai indikator fungsi ginjal. Pada kondisi fungsi ginjal normal, kreatinin dalam darah ada dalam jumlah konstan. Nilainya akan meningkat pada penurunan fungsi ginjal (Pedoman data interpretasi data klinik, 2011).

**Tabel 2.5** Nilai normal Kreatinin (*creatinine*)

Nilai Normal	0,6 – 1,3 mg/dL
--------------	-----------------

## 4. Berat Badan

Berat badan adalah ukuran yang lazim atau sering dipakai untuk menilai keadaan suatu gizi manusia, dengan mengetahui berat badan seseorang maka kita akan dapat memperkirakan tingkat kesehatan atau gizi seseorang.

## 5. Laju Filtrasi Glomerulus (LFG)

Laju Filtrasi Glomerulus merupakan pengukuran yang paling baik di dalam menilai fungsi ekskresi serta untuk mengetahui derajat penyakitir pada kerusakan pasien gagal ginjal kronis.

## 2.6 Data Mining

*Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait besar. *Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar. Dari definisi-definisi yang telah disampaikan ada beberapa bidang pekerjaan yang berkaitan dengan *data mining* menurut (Prasetyo, E, 2014) dibagi menjadi 4 (empat) kelompok, yaitu:

1. *Prediction Modelling* (model prediksi), pekerjaan ini berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat.
2. *Cluster Analysis* (analisis kelompok), melakukan pengelompokan data-data kedalam sejumlah kelompok berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada.

3. *Association Analysis* (analisis asosiasi), digunakan untuk menemukan pola yang menggambarkan kekuatan fitur dalam data dengan tujuan untuk menemukan pola yang menarik dengan cara yang efisien.
4. *Anomaly Detection* (deteksi anomali), berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain.

## 2.7 Proses Data Mining

Secara sistematis ada tiga langkah dalam data mining menurut (Gonunesco, 2011):

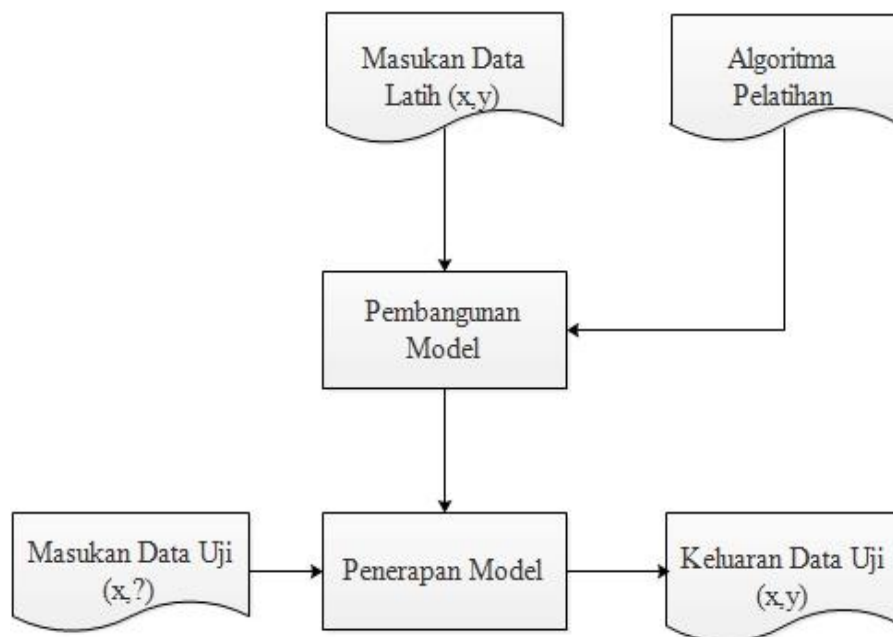
1. Eksplorasi/pemrosesan awal data  
Eksplorasi/pemrosesan awal data terdiri dari ‘pembersihan’ data, normalisasi data, transformasi data, penanganan data yang salah, reduksi dimensi, pemilihan subset fitur, dan sebagainya.
2. Membangun model dan melakukan validasi terhadapnya  
Membangun model dan melakukan validasi terhadapnya berarti melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang terbaik. Dalam langkah ini digunakan metode–metode seperti klasifikasi, regresi, analisis cluster, deteksi anomali, analisis asosiasi, analisis pola sekuensial, dan sebagainya.
3. Penerapan  
Penerapan berarti menerapkan model pada data yang baru untuk menghasilkan pekiraan/prediksi masalah yang diinvestigasi.

## 2.8 Klasifikasi

Menurut (Prasetyo, E, 2012), Klasifikasi dapat didefinisikan secara detail sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target  $f$  yang memetakan setiap vektor (set fitur)  $x$  ke satu dari sejumlah label kelas  $y$  yang tersedia. Model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, di mana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap



masukan tersebut, dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. Kerangka kerja (*framework*) klasifikasi ditunjukkan pada Gambar 2.1. Pada gambar tersebut, disediakan sejumlah data latih ( $x,y$ ) untuk digunakan sebagai data pembangunan model, kemudian menggunakan model tersebut untuk memprediksi kelas dari data uji ( $x,?$ ) sehingga data uji ( $x,?$ ) diketahui kelas  $y$  yang sesungguhnya.



**Gambar 2.1** Proses kerja klasifikasi

Model yang sudah dibangun pada saat pelatihan kemudian dapat digunakan untuk memprediksi label kelas data baru yang belum diketahui label kelasnya. Dalam pembangunan model selama proses pelatihan tersebut diperlukan adanya suatu algoritma untuk membangunnya yang disebut sebagai algoritma pelatihan (*learning algorithm*). Kerangka kerja seperti yang ditunjukkan pada **Gambar 2.1** meliputi dua langkah proses yaitu induksi dan deduksi. Induksi merupakan suatu langkah untuk membangun model klasifikasi dari data latih yang diberikan, disebut juga proses pelatihan. Sedangkan deduksi merupakan suatu langkah untuk menerapkan model tersebut pada data uji sehingga data uji dapat diketahui kelas yang sesungguhnya atau disebut juga proses prediksi.

## 2.9 Normalisasi Data

Normalisasi data adalah bertujuan untuk mempersempit atau mengecilkan nilai *range* pada data tersebut. Normalisasi yang digunakan pada penelitian ini adalah *min-max normalization* yang merupakan proses transformasi nilai dari data yang dikumpulkan pada *range value* antara 0.0 dan 1.0, dimana nilai terkecil (*min*) adalah 0.0 dan nilai tertinggi (*max*) adalah 1.0 (Chandrasekhar, Thangavel dan Elayaraja, 2011). Keuntungan dari metode ini adalah keseimbangan nilai perbandingan antara data saat sebelum dan sesudah nilai normalisasi. Kekurangannya adalah jika ada data baru metode ini akan memungkinkan terjebak pada *out of bound error*. Normalisasi data sangat di perlukan ketika data yang ada terlalu besar atau terlalu kecil sehingga pengguna kesulitan memahami informasi yang di maksud. Jika rentan nilai normalisasi yang di inginkan berada pada rentan [0,1] maka dapat juga menggunakan persamaan berikut :

$$\text{normalisasi } x_{ik} = \frac{x - \min(x_k)}{\max(x_k) - \min(x_k)} \dots\dots\dots (2.1)$$

Keterangan :

$x_{ik}$  = nilai hasil normalisasi

$x$  = nilai x sebelum normalisasi

$\min(x_k)$  = nilai minimum dari fitur

$\max(x_k)$  = nilai maksimum dari fitur

## 2.10 Metode *K-Nearest Neighbor* (KNN)

*K-Nearest Neighbors* (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. KNN termasuk dalam golongan *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. Nantinya kelas yang baru dari suatu data akan dipilih berdasarkan grup kelas yang paling dekat jarak vektornya.

Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru

berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah  $k$  obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari  $k$  obyek. Algoritma *K-Nearest Neighbors* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

Metode *K-Nearest Neighbors* sangatlah sederhana, bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN-nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai jika kelas  $c$  merupakan klasifikasi yang paling banyak ditemui pada  $k$  buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*.

Jarak *Euclidean* paling sering digunakan dalam menghitung jarak karena sangat cocok untuk menggunakan jarak terdekat (lurus) antara dua data. Jarak *euclidean* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek yang direpresentasikan sebagai berikut:

$$D(a, b) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \dots \dots \dots (2.2)$$

Keterangan :

- $D(a, b)$  = Jarak *Euclidean* Data  $a$  dan Data  $b$
- $X$  = Koordinat titik  $X$
- $Y$  = Koordinat titik  $Y$

Menurut Prasetyo (2012), salah satu masalah yang dihadapi KNN adalah pemilihan nilai  $K$  yang tepat. Misalnya, diambil  $K$  bernilai 13, kelas 0 dimiliki oleh 7 tetangga yang jauh, sedangkan kelas 1 dimiliki oleh 6 tetangga yang lebih dekat. Hal ini mengakibatkan data uji tersebut

akan terdistorsi sehingga ikut bergabung dengan kelas 0. Hal ini karena setiap tetangga tersebut mempunyai bobot yang sama terhadap data uji, sedangkan  $K$  yang terlalu kecil bisa menyebabkan algoritma terlalu sensitif terhadap *noise*. Nilai  $k$  yang bagus dapat dipilih berdasarkan optimisasi parameter, misalkan dengan *cross validation*. Kasus dimana klasifikasi diprediksikan berdasarkan *training* data yang paling dekat (dengan kata lain,  $k = 1$ ) disebut algoritma *nearest neighbors*.

Ketepatan algoritma KNN sangat dipengaruhi oleh ada atau tidaknya fitur- fitur yang tidak relevan atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur agar performa klasifikasi menjadi lebih baik. Berikut ini adalah urutan proses pada algoritma *K-Nearest Neighbors* :

1. Menentukan parameter  $K$  (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclidean* masing-masing obyek terhadap data sampel yang diberikan.
3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *euclidean* terkecil.
4. Mengumpulkan kategori  $Y$  (klasifikasi *nearest neighbors*).
5. Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai *query instance* yang telah dihitung.

KNN merupakan teknik klasifikasi sederhana, tetapi mempunyai hasil kerja yang cukup bagus. Meskipun begitu, KNN juga mempunyai kelebihan dan kekurangan. Beberapa karakteristik KNN adalah sebagai berikut (Eko Prasetyo, 2012) :

1. KNN merupakan algoritma yang menggunakan seluruh data latih untuk melakukan proses klasifikasi (*complete storage*). Hal ini mengakibatkan proses prediksi yang sangat lama untuk data dalam jumlah yang sangat besar. Pendekatan lain adalah dengan menggunakan *mean* data dari setiap kelas, kemudian menghitung jarak terdekat data uji ke *mean* data setiap kelas tersebut. Hal ini

memberi keuntungan kerja yang lebih cepat, tetapi hasilnya kurang memuaskan karena model hanya membentuk *hyperplane* tepat di tengah-tengah di antara 2 kelas yang memisahkan 2 kelas (untuk kasus 2 kelas). Semakin banyak data latih, semakin halus *hyperplane* yang dibuat. Ada relasi pertukaran (*trade-off relation*) antara jumlah data latih pada biaya komputasi dengan kualitas batas keputusan (*decision boundary*) yang dihasilkan.

2. Algoritma KNN tidak membedakan setiap fitur dengan suatu bobot seperti pada *Artificial Neural Network (ANN)* yang berusaha menekan fitur yang tidak mempunyai kontribusi terhadap klasifikasi menjadi 0 pada bagian bobot. KNN tidak memiliki bobot untuk masing-masing fitur.
3. Karena KNN masuk kategori *lazy learning* yang menyimpan sebagian atau semua data dan hampir tidak ada proses pelatihan, KNN sangat cepat dalam proses pelatihan (karena memang tidak ada), tetapi sangat lambat dalam proses prediksi.
4. Hal yang rumit adalah menentukan nilai K yang paling sesuai.
5. Karena KNN pada prinsipnya memilih tetangga terdekat, parameter jarak juga penting untuk dipertimbangkan sesuai dengan kasus datanya. *Euclidean* sangat cocok untuk menggunakan jarak terdekat

### 2.11 Metode Matriks Konfusi.

Sebuah sistem klasifikasi diharapkan dapat melakukan klasifikasi semua set data yang digunakan dengan benar, akan tetapi tidak dapat dihindari kinerja sebuah sistem tidak dapat 100% benar, maka dari itu dibutuhkan sebuah pengukuran kinerja sistem agar dapat mengetahui sejauh mana akurasi sistem yang telah dibuat. Matrik konfusi atau *confusion matrix* merupakan tabel pencatat hasil kerja klasifikasi, secara umum pengukuran akurasi perhitungan klasifikasi dilakukan

menggunakan matrik konfusi. **Tabel 2.6** merupakan contoh tabel matrik konfusi menggunakan 2 kelas.

**Tabel 2.6** Matriks konfusi

$f_{ij}$		Kelas hasil klasifikasi ( $j$ )	
		kelas = 1	kelas = 0
kelas asli ( $i$ )	kelas = 1	f11	f10
	kelas = 0	f01	f00

Keterangan :

$f_{ij}$  : Menyatakan setiap sel  $f_{ij}$  dalam matriks menyatakan jumlah data dari kelas ( $i$ ) yang hasil prediksinya masuk kelas ( $j$ ). Misalnya sel f11 adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, dan f10 adalah data dalam kelas 1 yang dipetakan salah ke kelas 0. Berdasarkan isi matrik konfusi dapat di ketahui jumlah data dari masing – masing kelas yang diprediksi benar yaitu ( $f_{11} + f_{00}$ ), dan data diklasifikasikan secara salah yaitu ( $f_{10} + f_{01}$ ). Kuantitas matrik konfusi dapat diringkas menjadi dua nilai yaitu akurasi dan laju error Untuk mengukur nilai akurasi yang didapat dari hasil pengujian menggunakan rumus 2.3.

$$Akurasi = \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah prediksi yang dilakukan}} \\ \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \times 100\% \quad \dots \dots \dots (2.3)$$

Sedangkan untuk mengukur tingkat kesalahannya menggunakan rumus 2.4.

$$Laju\ error = \frac{\text{jumlah data yang diprediksi salah}}{\text{jumlah prediksi yang dilakukan}} \\ \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \times 100\% \quad \dots \dots \dots (2.4)$$

## 2.12 Penelitian Sebelumnya

Penulis mengkaji dari hasil–hasil penelitian yang memiliki kesamaan topik dengan yang sedang diteliti oleh penulis. Adapun beberapa kajian yang berhubungan dengan topik yang sedang diteliti, antara lain:

1. Mei Lestari, “*Penerapan Algoritma Klasifikasi Nearest Neighbor (KNN) Untuk Mendeteksi Penyakit Jantung*”. Tahun 2014, Universitas Indraprasta Jakarta Selatan. Dalam penelitian ini dilakukan penerapan algoritma K-NN dengan  $k = 9$  pada data pasien untuk mendeteksi penyakit jantung. Kedekatan antara kasus pada data training dan data uji dilakukan untuk menentukan kelas data uji tersebut. Untuk mengukur kinerja algoritma tersebut dilakukan dengan menggunakan confusion matrix dan kurva ROC, diperoleh nilai akurasi 70% dan termasuk klasifikasi baik karena memiliki nilai AUC 0.875. sehingga algoritma K-NN dapat digunakan dan diterapkan untuk mendeteksi penyakit jantung.
2. Andi Sri Irtawaty, “*Klasifikasi Penyakit Ginjal Dengan Metode K-Means*”. Tahun 2017, Fakultas Teknik Elektronika, Politeknik Negeri Balikpapan. Pada penelitian ini mengklasifikasikan keluaran penyakit gagal ginjal dengan menggunakan menggunakan metode *K-Means*. Data yang digunakan adalah data uji berjumlah 10 data pasien gagal ginjal. Hasil penelitian yang dilakukan dapat digunakan untuk mengetahui hasil keluaran stadium normal, satu, dua, tiga dan empat. Hasil dari pengujian yang di dapat selanjutnya di bandingkan dengan data literatur, studi literatur diperoleh dari buku, jurnal dan artikel laporan penelitian.