

BAB II

LANDASAN TEORI

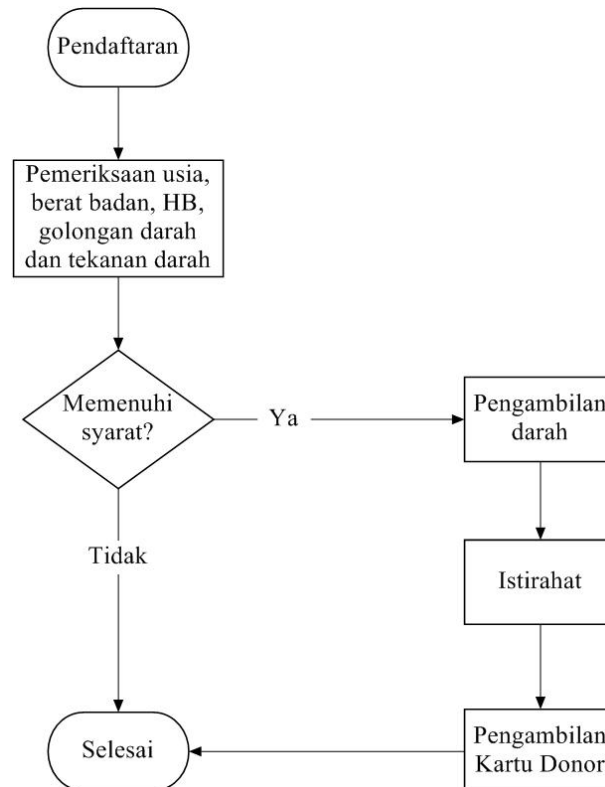
2.1 Donor Darah

Donor darah merupakan proses pengambilan darah dari seseorang secara sukarela untuk disimpan di bank darah untuk kemudian dipakai pada transfusi darah bagi pasien yang membutuhkan. Untuk dapat menyumbangkan darah, seorang donor darah harus memenuhi syarat sebagai berikut :

1. Berbadan sehat
2. Usia 17-60 tahun (pada usia 17 tahun diperbolehkan menjadi donor bila mendapat ijin tertulis dari orang tua. Sampai usia tahun donor masih dapat menyumbangkan darahnya dengan jarak penyumbangan 3 bulan atas pertimbangan dokter).
3. Berat badan minimum 45 kg.
4. Temperatur tubuh : 36,6 - 37,5 °C (oral).
5. Tekanan darah baik, yaitu:
 - a. Sistole : 110 - 160 mm Hg.
 - b. Diastole : 60 - 100 mm Hg.
6. Denyut nadi : Teratur 50 - 100 kali/ menit.
7. Haemoglobin
 - a. Wanita : minimal 12 gr %
 - b. Pria : minimal 12,5 gr %
8. Jumlah penyumbangan pertahun paling banyak 4 kali dengan jarak penyumbangan sekurang-kurangnya 3 bulan. Keadaan ini harus sesuai dengan keadaan umum donor.
9. Bagi penyumbang darah wanita tidak sedang menstruasi, hamil atau menyusui.
10. Tidak dalam pengaruh obat-obatan seperti golongan narkotika dan alkohol.
11. Tidak menderita penyakit: jantung, hati, paru-paru, ginjal, kencing manis, penyakit kelainan darah, gangguan pembekuan darah, epilepsi, kanker atau penyakit kulit.

2.2 Prosedur Donor Darah

Sebelum dilakukan proses pengambilan darah, calon pendonor darah akan melalui beberapa tahap. Berikut ini akan dijelaskan secara lebih detail alur donor darah menggunakan *flowchart* yang disajikan pada **gambar 2.1**.



Gambar 2.1 *Flowchart* prosedur donor darah

Tahap pertama bagi calon pendonor darah yaitu mengambil *form* pendaftaran dan mengisi data pribadi. Kemudian dilakukan pemeriksaan usia, berat badan, kadar HB, golongan darah dan tekanan darah oleh petugas UDD PMI. Jika memenuhi syarat, tahap berikutnya yaitu pengambilan darah. Sedangkan bagi yang tidak memenuhi syarat donor darah, maka pengambilan darah tidak dapat dilakukan. Setelah selesai, pendonor darah diberi waktu beristirahat sejenak. Apabila keadaan sudah membaik, petugas UDD PMI mempersilahkan pendonor darah untuk mengambil kartu donor. Kartu tersebut adalah bukti bahwa seseorang telah melakukan donor darah. Proses donor darahpun dinyatakan selesai.

2.3 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:

- a. Komponen sistem (components)
- b. Batas sistem (boundary)
- c. Lingkungan luar sistem (environments)
- d. Penghubung sistem (interface)
- e. Masukan sistem (input)
- f. Keluaran sistem (output)
- g. Pengolah sistem
- h. Sasaran sistem (objective)

2.4 Data Mining

Secara sederhana, *data mining* merupakan ekstraksi informasi yang tersirat dalam sekumpulan data. Data mining merupakan sebuah proses untuk menggali kumpulan data dan menemukan informasi di dalamnya. (Turban, E., dkk. 2005). Data mining merupakan proses pengekstrakan informasi dari

jumlah kumpulan data yang besar dengan menggunakan algoritma dan teknik gambar dari statistik, mesin pembelajaran dan sistem manajemen *database*. Penggalian data ini dilakukan pada sekumpulan data yang besar untuk menemukan pola atau hubungan yang ada dalam kumpulan data tersebut (Kusrini dan E.T. Lutfi. 2009). Hasil penemuan yang diperoleh setelah proses penggalian data ini, kemudian dapat digunakan untuk analisis yang lebih lanjut.

Data mining yang disebut juga dengan *Knowledge-Discovery in Database* (KDD) adalah sebuah proses secara otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi, hubungan (*association*) atau pengelompokan (*clustering*). Proses KDD ini terdiri dari langkah-langkah sebagai berikut (Han, J. dan M. Kamber. 2006):

1. *Data Cleaning*, proses menghapus data yang tidak konsisten dan kotor.
2. *Data Integration*, penggabungan beberapa sumber data.
3. *Data Selection*, pengambilan data yang akan dipakai dari sumber data.
4. *Data Transformation*, proses dimana data ditransformasikan menjadi bentuk yang sesuai untuk diproses dalam data mining.
5. *Data Mining*, suatu proses yang penting dengan melibatkan metode untuk menghasilkan suatu pola data.
6. *Pattern Evaluation*, proses untuk menguji kebenaran dari pola data yang mewakili *knowledge* yang ada didalam data itu sendiri.
7. *Knowledge Presentation*, proses visualisasi dan teknik menyajikan *knowledge* digunakan untuk menampilkan *knowledge* hasil *mining* kepada *user*.

2.5 Metode Data Mining

Pada umumnya metode *data mining* dapat dikelompokkan kedalam dua kategori yaitu *deskriptif* dan *prediktif*. Metode *deskriptif* bertujuan untuk mencari pola yang dapat dimengeti oleh manusia yang menjelaskan karakteristik dari data. Metode *prediktif* menggunakan ciri-ciri tertentu dari

data. Pada umumnya metode *data mining* dapat dikelompokkan kedalam dua untuk melakukan prediksi.

Metode-metode yang ada dalam *data mining* adalah sebagai berikut (Tang, ZhaoHui and J. MacLennan. 2005):

1. *Classification*

Klasifikasi (*Classification*) merupakan proses untuk menemukan sekumpulan model yang menjelaskan dan membedakan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah objek. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih (*training set*). Sedangkan data uji (*test set*) digunakan untuk mengetahui tingkat akurasi dari model yang dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu obyek data. Metode inilah yang digunakan dalam tugas akhir ini.

2. *Clustering*

Pengelompokan (*Clustering*) merupakan proses untuk melakukan segmentasi. Digunakan untuk melakukan pengelompokan secara alami terhadap atribut suatu set data, termasuk kedalam *supervised task*. Contoh *clustering* seperti mengelompokkan dokumen berdasarkan topiknya.

3. *Assosiation*

Tujuan dari metode ini untuk menghasilkan sejumlah *rule* yang menjelaskan sejumlah data yang berhubung kuat satu dengan yang lainnya. Sebagai contoh *assosiation analysis* dapat digunakan untuk menentukan produk yang datang secara bersamaan oleh banyak pelanggan, atau bisa juga disebut dengan *basket analysis*.

4. *Regression*

Regression mirip dengan klasifikasi. Perbedaan utamanya adalah terletak pada atribut yang diprediksi berupa nilai yang kontinyu.

5. *Forecasting*

Prediksi (*Forecasting*) berfungsi untuk melakukan kejadian yang akan datang berdasarkan data sejarah yang ada.

6. *Sequence Analysis*

Tujuan dari metode ini adalah untuk mengenali pola dari data diskrit. Sebagai contoh adalah menemukan kelompok gen dengan tingkat ekpresi yang mirip.

7. *Deviation Analysis*

Tujuan dari metode ini adalah untuk menemukan penyebab perbedaan antara data yang satu dengan data yang lain dan biasa disebut sebagai *oulier detection*. Sebagai contoh adalah apakah sudah terjadi penipuan terhadap pengguna kredit dengan melihat catatan transaksi yang tersimpan dalam basis data perusahaan kartu kredit.

2.6 Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan/klasifikasi /prediksi pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpannya. Model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, dimana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. (Prasetyo, E. 2012).

Tahapan dari klasifikasi dalam data mining terdiri dari (Han, J. dan M. Kamber. 2006) :

1. Pembangunan Model

Pada tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi class atau atribut dalam data. Tahap ini merupakan fase pelatihan, dimana data latih dianalisis menggunakan algoritma klasifikasi, sehingga model pembelajaran direpresentasikan dalam bentuk aturan klasifikasi.

2. Penerapan Model

Pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan atribut/kelas dari sebuah data baru yang atribut/kelasnya belum diketahui sebelumnya. Tahap ini digunakan untuk memperkirakan keakuratan aturan klasifikasi terhadap data uji. Jika model dapat diterima, maka aturan dapat diterapkan terhadap klasifikasi data baru.

2.7 *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 :

$$D(a, b) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \dots\dots\dots (2.1)$$

Keterangan :

$D(a, b)$ = Jarak Euclidean Data a dan Data b

X = Koordinat titik X

Y = Koordinat titik Y

Menurut Eko Prasetyo (Eko Prasetyo, 2012) dalam buku Konsep dan Aplikasi *Data Mining*, 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 tergabung 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 (Aryo, 2012). 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.

2.7.1 Normalisasi Data

Normalisasi data linier adalah proses penskalaan nilai atribut data sehingga bisa jatuh pada range tertentu. Tujuan dari normalisasi data adalah 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 :

$$\text{normalisasi } X^* = \frac{(X - \min X)}{(\max X - \min X)} \dots\dots\dots(2.2)$$

Keterangan :

- X^* = nilai hasil normalisasi
- X = nilai x sebelum normalisasi,
- \min = nilai minimum dari fitur,
- \max = nilai maksimum dari fitur,

2.7.2 Flowchart Perhitungan *K- Nearest Neighbor*



Gambar 2.2 Flowchart *K-Neraest Neighbor*

dengan d adalah jarak antara titik pada data training x dan titik data testing y yang akan diklasifikasi, dimana $x=x_1,x_2,\dots,x_i$ dan $y=y_1,y_2,\dots,y_i$ dan I merepresentasikan nilai atribut serta n merupakan dimensi atribut (Han & Kamber, 2001). Sebagai ilustrasi, pada Tabel 2.1 berikut ini disajikan contoh penerapan rumus *Euclidean*, pada empat data klasifikasi kualitas baik dan tidak baik sebuah kertas tisu yang dinilai berdasarkan daya tahan kertas tersebut dan fungsinya. Sebanyak tiga data yang sudah terklasifikasi yaitu data no 1,2, dan 3. masing-

masing data dihitung jaraknya ke data no 4 untuk mendapatkan kelas yang sesuai bagi data no 4 maka k=1 klasifikasi kualitas baik atau tidak baik sebuah kertas tisu

Tabel 2.1 Set data pada contoh metode K-NN

No	Fungsi	Daya Tahan	Klasifikasi
1	7	7	Tidak Baik
2	7	4	Tidak Baik
3	3	4	Baik
4	1	4	?

Berikut ini disajikan pula perhitungan yang dilakukan terhadap tiga data yang sudah terklasifikasi dengan data yang belum terklasifikasi pada Tabel di atas. Jarak data no satu ke data no empat:

$$d_{1,4} = \sqrt{(7 - 1)^2 + (7 - 4)^2} = \sqrt{6^2 + 3^2} = \sqrt{45} = 6.07$$

Jarak data no dua ke data no empat:

$$d_{2,4} = \sqrt{(7 - 1)^2 + (4 - 4)^2} = \sqrt{6^2 + 0^2} = \sqrt{36} = 6$$

Jarak data no tiga ke data no empat:

$$d_{3,4} = \sqrt{(3 - 1)^2 + (4 - 4)^2} = \sqrt{2^2 + 0^2} = \sqrt{4} = 2$$

Dari hasil perhitungan di atas diperoleh jarak antara data no tiga dan data no empat adalah jarak yang terdekat maka kelas data no empat adalah baik. Teknik ini akan diuji cobakan terhadap dataset yang belum terklasifikasi atau data yang belum dikenal, untuk menemukan kelas yang sesuai dengan berdasarkan pada data tetangga terdekatnya yang sudah terklasifikasi. Tingkat ketepatan klasifikasi terhadap data dari kedua algoritma yang digunakan menjadi titik fokus analisa dalam penelitian.

2.8 Evaluasi Sistem

Menurut (Olson, David, Yong Shi. 2008) tujuan evaluasi percobaan pada klasifikasi yaitu untuk mengukur keefektifan apakah sistem mengklasifikasi secara benar. Pada pengukuran kesalahan biasanya membutuhkan sebuah

matriks yang disebut berupa *confusion matrix*. Berikut merupakan tabel *confusion matrix*.

Tabel 2.2 *confusion matrix*

Kelas Hasil Prediksi Kelas Asli	Positif	Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Keterangan **Tabel 2.1** :

1. TP adalah jumlah dari kelas positif (kelas yang mempunyai jumlah data lebih sedikit) yang benar diklasifikasikan.
2. FN adalah jumlah kelas positif yang salah diklasifikasikan kedalam kelas negatif.
3. FP adalah jumlah kelas negatif yang salah diklasifikasikan kedalam kelas positif.
4. TN adalah jumlah kelas negatif yang benar diklasifikasikan.

Berdasarkan data yang didapatkan dari *confusion matrix*, maka akan dihitung (Weng dan Poon 2008) :

1. Akurasi dan Laju Error

Akurasi digunakan untuk mengukur prosentase pengenalan secara keseluruhan dan dihitung sebagai jumlah data uji yang dikenali dengan benar, dibagi dengan jumlah seluruh data uji. Berikut rumus akurasi dan laju *error* pada persamaan 2.3 dan persamaan 2.4.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} \\
 &= \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad \dots \dots \dots (2.3)
 \end{aligned}$$

$$\begin{aligned} \text{Laju error} &= \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{Jumlah prediksi yang dilakukan}} \\ &= \frac{FN+FP}{TP+FN+FP+TN} \times 100\% \quad \dots \dots \dots (2.4) \end{aligned}$$

Sensitivitas akan mengukur proporsi positif asli yang dikenali (diprediksi) secara benar sebagai positif asli. Rumus perhitungannya menggunakan rumus 2.5. Sedangkan spesifisitas akan mengukur proporsi negatif asli yang dikenali (diprediksi) secara benar sebagai negatif asli. Rumus perhitungannya menggunakan rumus 2.6.

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \times 100\% \quad \dots \dots \dots (2.5)$$

$$\text{Spesifisitas} = \frac{TN}{FP + TN} \times 100\% \quad \dots \dots \dots (2.6)$$

2.9 Penelitian yang Terkait

Penelitian sebelumnya dilakukan oleh seorang mahasiswa Universitas Muhammadiyah Gresik yang bernama Lailatul Qomariyah dengan judul “*Klasifikasi Calon Pendorong Darah dengan Metode Decision Tree C4.5 di Kabupaten Gresik (Studi Kasus : PMI KAB. GRESIK)*”. Penelitian ini membahas tentang klasifikasi calon pendonor darah berdasarkan variabel-variabel yang telah ditentukan dengan menggunakan metode decision tree c4.5. Hasil klasifikasi dari sistem ini akan digunakan untuk menentukan calon pendonor darah yang layak mendonorkan darahnya sesuai syarat-syarat yang telah ditentukan. Sehingga hasil pengujian diketahui bahwa akurasi tertinggi yang diperoleh 91,67% pada pengujian ketiga.

Selain itu metode *KNN* juga digunakan untuk penelitian seseorang yang bernama Sumarlin mahasiswa Universitas STIKOM Uyelindo, Kupang dengan judul “*Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM*”. Penelitian ini membahas tentang klasifikasi beasiswa Peningkatan Prestasi Akademik (PPA)

dan Bantuan Belajar Mahasiswa (BBM) berdasarkan variabel-variabel yang telah ditentukan dengan menerapkan algoritma *k-nearest neighbor*. Hasil klasifikasi dari sistem ini akan digunakan sebagai keputusan dalam pemberian beasiswa bagi mahasiswa yang mengajukannya. Sehingga dari hasil pengujian diketahui bahwa akurasi yang diperoleh untuk beasiswa Peningkatan Prestasi Akademik (PPA) mencapai 83.33% dan beasiswa Bantuan Belajar Mahasiswa (BBM) mencapai 90%.

Sri Redjeki, jurusan Teknik Informatika, STMIK AKAKOM. Penelitiannya yang berjudul "*Identifikasi Penyakit dengan Gejala Awal Demam Menggunakan K-Nearest Neighbor (K-NN)*". Pada penelitian ini menggunakan data pasien dengan gejala awal demam untuk mengelompokkan penyakit yang terdiri dari 10 penyakit. Obyek penelitian menggunakan data sebanyak 82 dengan 72 data digunakan untuk training dan 10 data digunakan untuk testing. Hasil terbaik pengelompokan penyakit menggunakan nilai $K=3$ dengan nilai akurasi hasil pengelompokan sebesar 97,2%. Hal ini menunjukkan bahwa metode K-NN merupakan bagian dari CDSS karena nilai akurasi yang dapat ditoleransi untuk pengelompokan penyakit harus mempunyai nilai akurasi diatas 97%.

Sisca Lidhya Sari, Universitas Sumatera Utara. Penelitiannya yang berjudul "*Implementasi Algoritma K-Nearest Neighbor Dan Metode Topsis (Technique For Orders Preference By Similarity To Ideal Solution) Dalam Penentuan Mutu Beras Miskin (Studi Kasus: Bulog Aceh)*" kriteria-kriteria yang digunakan dalam penelitian ini yaitu : Bebas Hama Penyakit, Butir Patah, Kadar Air, Bebas Bahan Kimia. Hasil dari penelitian tersebut, Sistem dapat dapat digunakan untuk menentukan mutu beras. Kesimpulan dari pengujian yang telah dilakukan, maka antara K-Nearest Neighbor dan TOPSIS terlihat algoritma K-Nearest Neighbor lebih sederhana proses perhitungannya terhadap data beras dibanding dengan Metode Topsis.