

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:

- a. Komponen Sistem
- b. Batas Sistem
- c. Lingkungan luar Sistem
- d. Pembhubung Sistem
- e. Masukan Sistem
- f. Keluaran Sistem
- g. Pengolahan Sistem
- h. Sasaran sistem

2.2 Stroke

Stroke merupakan penyakit yang terjadi akibat penyumbatan pada pembuluh darah otak atau pecahnya pembuluh darah di otak yang mendadak. Sehingga akibat penyumbatan maupun pecahnya pembuluh darah tersebut, bagian otak tertentu berkurang bahkan terhenti suplai oksigennya sehingga menjadi rusak bahkan mati. Akibatnya timbulah berbagai macam gejala sesuai dengan daerah otak yang terlibat, seperti wajah lumpuh sebelah, bicara pelo (cedal), lumpuh anggota gerak, bahkan sampai koma dan dapat mengancam jiwa.

2.3 Penjelasan Stadium Stroke

1. Stadium Stroke Rendah

Stroke ringan dapat menyebabkan kelemahan pada otot wajah, tandatanya adalah wajah turun ke satu sisi (wajah terlihat tidak simetris), tidak dapat mengerutkan dahi, dan mata atau mulut turun kebawah.

2. Stadium Stroke Sedang

Penderita Stroke rendah mengalami kelumpuhan pada salah satu sisi tubuh, seperti lengan atau kaki. Kesulitan berjalan atau sulit mempertahankan posisi tubuh karena adanya gangguan system koordinasi tubuh.

3. Stadium Stroke Tinggi

Seseorang dengan stadium penyakit stroke tinggi tidak dapat berjalan maupun menggerakkan bagian tangan, dan penderita mengalami kelumpuhan hampir 80% pada bagian tubuhnya.

2.4 Jenis - Jenis Stroke

Gangguan peredaran darah di otak dapat menyebabkan stroke. Namun penyakit stroke masih digolongkan menjadi dua jenis yaitu stroke iskemik, yang disebabkan karena penyumbatan pembuluh darah otak dan stroke hemoragik disebabkan oleh pecahnya pembuluh darah otak (Yastroki, 2011).

1. Stroke Iskemik

Proses stroke yang disebabkan oleh penyumbatan pembuluh darah otak yang menyebabkan berkurangnya suplai oksigen dan glukosa ke otak. Jaringan otak akan mati jika suplai oksigen secara berkelanjutan terus berkurang, sehingga pembekuan darah dapat terjadi.

2. Stroke Hemoragik

Selain karena penyumbatan, pendarahan didalam otak dapat terjadi sehingga mengganggu aliran darah ke otak. Stroke hemoragik menjadi lebih berbahaya dan menyebabkan kematian. Pendarahan yang terjadi pada stroke hemoragik dibagi menjadi dua jenis dengan pendarahan yang terjadi didalam otak dan pendarahan diantara bagian dalam dan luar lapisan jaringan yang melindungi otak.

2.5 Faktor Risiko Penyakit Stroke

Faktor risiko stroke adalah kondisi atau penyakit atau kelainan yang terdapat pada seseorang yang memiliki potensi untuk memudahkan orang tersebut mengalami serangan stroke pada suatu saat. Jika seseorang terdapat faktor-faktor risiko untuk terjadinya serangan stroke disebut sebagai *stroke prone profile* (Kemenkes, 2013).

Terdapat dua macam faktor risiko penyakit stroke, yaitu faktor risiko yang dapat diubah atau dikendalikan dan faktor risiko yang tidak dapat diubah atau dikendalikan. Faktor risiko yang tidak dapat diubah atau dikendalikan (Valensia, S.A. 2015) meliputi :

1. Usia

Stroke sering terjadi pada orang yang telah lanjut usia (tua). Setiap penambahan 10 tahun setelah usia 55 tahun, terdapat peningkatan risiko penyakit stroke sebanyak dua kali lipat.

2. Jenis Kelamin

Stroke lebih mungkin pada pria dibandingkan pada wanita. Namun, lebih dari separuh kematian stroke total yang terjadi pada wanita.

Penggunaan pil KB dan kehamilan meningkatkan risiko stroke bagi perempuan.

3. Ras

Kematian akibat penyakit stroke lebih banyak terjadi pada orang Afrika-Amerika daripada orang kulit putih. Hal ini dikarenakan mereka mempunyai risiko lebih tinggi menderita tekanan darah tinggi, diabetes, dan obesitas.

Sedangkan faktor risiko yang dapat diubah atau dikendalikan meliputi:

4. Tekanan Darah

Tekanan darah adalah tekanan yang terjadi pada pembuluh darah arteri saat darah dipompa oleh jantung untuk dialirkan ke seluruh tubuh. Tekanan darah sistolik adalah tekanan darah yang terjadi pada saat otot jantung berkontraksi. Sedangkan tekanan darah diastolik adalah tekanan darah yang terjadi pada saat otot jantung beristirahat atau tidak sedang berkontraksi

5. Kadar Gula Darah

Gula darah adalah bahan bakar tubuh yang dibutuhkan untuk kerja otak, sistem saraf, dan jaringan tubuh yang lain. Gula darah yang terdapat di dalam tubuh dihasilkan oleh makanan yang mengandung karbohidrat, protein, dan lemak

6. Kadar Kolesterol Total

Kolesterol total merupakan kadar keseluruhan kolesterol yang beredar dalam tubuh manusia. Kolesterol adalah lipid amfipatik dan merupakan komponen struktural esensial pada membran plasma. Senyawa kolesterol total ini disintesis di banyak jaringan dari asetil-KoA dan merupakan prekursor utama semua steroid lain di dalam tubuh termasuk kortikosteroid, hormone seks, asam empedu, dan vitamin D.

7. *Low Density Lipoprotein (LDL)*

Kolesterol LDL disebut sebagai kolesterol jahat disebabkan peranannya membawa kolesterol total ke banyak jaringan di dalam

tubuh. Sehingga memberikan peluang terjadinya penumpukan kolesterol di berbagai jaringan tubuh, termasuk diantaranya dalam pembuluh darah.

8. Asam Urat

Penyakit asam urat adalah penyakit yang timbul akibat kadar asam urat darah yang berlebihan. Yang menyebabkan kadar asam urat darah berlebihan adalah produksi asam urat di dalam tubuh lebih banyak dari pembuangannya. Organ yang bisa terserang adalah sendi, otot, jaringan di sekitar sendi, telinga, kelopak mata, jantung, ginjal, dan lain-lain.

9. Kreatinin (*creatinine*)

Kreatinin (*creatinine*) adalah produk penguraian dari kreatin fosfat dalam metabolisme otot dan dihasilkan dari kreatin (*creatine*). Kreatinin pada dasarnya merupakan limbah kimia yang selanjutnya diangkut ke ginjal melalui aliran darah untuk dikeluarkan melalui urin. Kadar kreatinin dapat diukur dalam urin serta darah. Tingkat kreatinin dalam darah umumnya tetap normal karena massa otot relatif konstan. Dengan demikian, ginjal yang berfungsi normal juga akan menunjukkan tingkat normal kreatinin dalam darah. Tapi ketika ginjal tidak berfungsi dengan baik, jumlah kreatinin dalam darah akan meningkat.

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 dibagi menjadi 4 (empat) kelompok, yaitu (Prasetyo, E, 2012) :

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 **Klasifikasi**

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu.

Tahapan dari klasifikasi dalam data mining menurut (Han dan Kamber, 2006) terdiri dari :

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/class dari sebuah data baru yang atribut/classnya 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.

2.8 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. 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

2.9 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 obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari 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*. 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 (Data Uji Setiap Variabel)

Y = Koordinat titik Y (Data Latih Setiap Variabel)

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 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. 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.

2.10 Metode Pengukuran Kesalahan

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 2.1 *Confusion Matrix*

Kelas Asli \ Kelas Hasil Prediksi	Positif	Negatif
	Positif	True Positive (TP)
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 :

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.

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

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

2.11 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:

Penelitian sebelumnya dilakukan oleh seorang mahasiswa Universitas Muhammadiyah Gresik yang bernama Sugarwanto Atmaja untuk skripsi yang berjudul “**Sistem Pendukung Keputusan Untuk Deteksi Dini Risiko Penyakit Stroke Menggunakan *Learning Vector Quantization***”. Pada penelitian ini dilakukan 4 pengujian dengan pengujian pertama menggunakan 90 data latih dan 38 data uji dengan parameter laju pembelajaran 0.04 ; fungsi pembelajaran 0.6 ; jumlah iterasi 7 ; minal alfa 0.001 dan bobot yang digunakan 0.5. Selanjutnya pada pengujian ke dua menggunakan 90 data latih dan 38 data uji dengan parameter laju pembelajaran 0.05 ; fungsi pembelajaran 0.6 ; jumlah iterasi 7 ; minal alfa 0.001 dan bobot yang digunakan 0.5. Sedangkan untuk pengujian ke tiga menggunakan 90 data latih dan 38 data uji dengan parameter laju pembelajaran 0.04 ; fungsi pembelajaran 0.6 ; jumlah iterasi

7 ; minal alfa 0.001 dan bobot yang digunakan 0.5. Dan yang terakhir yaitu pada pengujian ke empat menggunakan 90 data latih dan 38 data uji dengan parameter laju pembelajaran 0.05 ; fungsi pembelajaran 0.6 ; jumlah iterasi 7 ; minal alfa 0.001 dan bobot yang digunakan 0.5. Dari hasil pengujian pertama sampai dengan pengujian ke empat dapat disimpulkan bahwa pada pengujian pertama sampai dengan pengujian ke tiga memiliki nilai akurasi yang sama yaitu sebesar 82%, sedangkan untuk pengujian ke empat mendapatkan nilai akurasi sebesar 79%.

Selain itu metode *K-Nearest Neighbor* juga digunakan untuk penelitian seseorang yang bernama Mei Lestari mahasiswa Universitas Indraprasta, Jakarta Selatan dengan judul “**Penerapan Algoritma Klasifikasi Nearest Neighbor (KNN) Untuk Mendeteksi Penyakit Jantung**”. Penelitian ini membahas tentang penerapan algoritma *K-Nearest Neighbor* untuk mendeteksi penyakit jantung. Pada penelitian ini digunakan algoritma K-NN dengan nilai K=9 pada 100 data pasien penyakit jantung. Hasil penelitian diperoleh nilai akurasi sebesar 70% serta nilai *AUC* (area dibawah kurva) sebesar 0.875 yang masuk kedalam klasifikasi baik, sehingga algoritma K-NN dapat digunakan dan diterapkan untuk mendeteksi penyakit jantung.