

BAB II

LANDASAN TEORI

2.1 Stroke

Stroke adalah serangan otak yang timbul secara mendadak dimana terjadi gangguan fungsi otak sebagian atau menyeluruh sebagai akibat dari gangguan aliran darah oleh karena sumbatan atau pecahnya pembuluh darah tertentu di otak, sehingga menyebabkan sel-sel otak kekurangan darah, oksigen atau zat-zat makanan dan akhirnya dapat terjadi kematian sel-sel tersebut dalam waktu relatif singkat (Yastroki, 2011).

Stroke merupakan gangguan fungsi saraf pusat yang berkembang sangat cepat baik menit maupun jam dengan perburukan ringan sampai berat kemudian menetap atau bahkan membaik secara cepat atau perlahan-lahan tergantung tingkat keparahan stroke dan cepat serta tepatnya intervensi pengobatan. Karena setiap bagian otak memiliki fungsi-fungsi tertentu, maka gejala dan tanda stroke pada setiap individu sangat bervariasi, tergantung pembuluh darah mana yang terkena dan bagian otak mana yang terganggu (Yastroki, 2011).

2.2 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 (Yastroki, 2011). Jika seseorang terdapat faktor-faktor risiko untuk terjadinya serangan stroke disebut sebagai *stroke prone profile*.

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 (Valencia, S.A. 2015) meliputi:

1. 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

2. 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. Rata-rata, kadar gula darah normal adalah sebagai berikut:

- a. Gula darah 8 jam sebelum makan atau setelah bangun pagi (70-110 mg/dl).
- b. Gula darah 2 jam setelah makan (100-150 mg/dl).
- c. Gula darah acak (70-125 mg/dl).

3. 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.

4. *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.

5. 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.

6. *Blood Urea Nitrogen*(BUN)

Blood Urea Nitrogen (BUN) dapat didefinisikan sebagai jumlah nitrogen urea yang hadir dalam darah. Urea adalah produk limbah yang dibentuk dalam tubuh selama proses pemecahan protein. Selama metabolisme protein, protein diubah menjadi asam amino yang juga menghasilkan amonia. Urea tidak lain adalah substansi yang dibentuk oleh beberapa molekul amonia. Metabolisme protein berlangsung dalam hati dan dengan demikian urea juga diproduksi oleh hati. Selanjutnya, urea ditransfer ke ginjal melalui aliran darah dan dikeluarkan dari tubuh dalam

bentuk urin. Dengan demikian, setiap disfungsi ginjal akan menyebabkan kadar tinggi atau rendah BUN dalam darah.

7. 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.

2.3 Pengendalian Stroke

Untuk menurunkan angka kesakitan, kecacatan dan kematian diperlukan pengendalian stroke. Kegiatan pengendalian stroke meliputi :

1. Pelayanan pra stroke

Pelayanan pra stroke adalah kegiatan deteksi dini, penemuan dan monitoring faktor risiko stroke pada individu sehat dan berisiko di masyarakat. Pelayanan pra stroke dilakukan di:

- a. Puskesmas
- b. Klinik kesehatan
- c. Posbindu PTM

2. Pelayanan serangan stroke

Pelayanan serangan stroke dilakukan di:

- a. Rumah sakit dipusatkan pada unit stroke atau pojok stroke
- b. Rumah sakit khusus

3. Pelayanan paska stroke

Pelayanan paska stroke dilakukan di:

- a. Rumah sakit
- b. Puskesmas

c. Posbindu PTM

2.4 Definisi Sistem

Sistem secara fisik adalah kumpulan dari elemen-elemen yang beroperasi bersama-sama untuk menyelesaikan suatu sasaran (Gordon, 1991). Menurut (Jagianto, 2005) Sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu.

2.5 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. Model itu sendiri bisa berupa aturan "Jika-maka", berupa pohon keputusan, atau formula matematis. (Bustami, 2011)

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 attribut 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 attribut/class dari sebuah data baru yang attribut/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.

2.6 Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, Daniel T. 2005).

1. Deskripsi

Deskripsi adalah menggambarkan pola dan kecenderungan yang terdapat dalam data secara sederhana. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Klasifikasi menggunakan *supervised learning*.

3. Estimasi

Estimasi hampir sama dengan klasifikasi, perbedaannya adalah variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun dengan menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

4. Prediksi

Prediksi memiliki kesamaan dengan klasifikasi dan estimasi, perbedaannya adalah hasil dari prediksi akan ada dimasa mendatang. Beberapa teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk prediksi.

5. Klastering

Klastering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Klastering menggunakan *unsupervised learning*.

6. Asosiasi

Tugas asosiasi atau sering disebut juga sebagai *market basket analysis* dalam data mining adalah menemukan relasi atau korelasi diantara himpunan item-item dan menemukan atribut yang muncul dalam satu waktu. Asosiasi menggunakan *unsupervised learning*. Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, *support* dan *confidence*.

Metode yang akan digunakan pada penelitian ini termasuk kedalam kelompok prediksi, karena menggunakan teknik klasifikasi yang hasilnya akan ada dimasa mendatang.

2.7 Normalisasi Data Linear

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

$$\text{normalisasi } (X_{ik}) = \frac{X_{ik} - \min(X_k)}{\max(X_k) - \min(X_k)} \dots \dots \dots (2.1)$$

X= nilai yang akan dinormalisasi,

k = kolom kriteria,

i = baris data,

min = nilai terkecil dari semua data i,

max = nilai terbesar dari semua data i,

2.8 FK-NN (*Fuzzy K-Nearest Neighbor*)

Selain K-NN yang melakukan prediksi secara tegas pada uji berdasarkan perbandingan K tetangga terdekat, ada pendekatan lain yang dalam melakukan

prediksi juga berdasarkan K tetangga terdekat, tetapi tidak secara tegas memprediksi kelas yang harus diikuti oleh data uji, pemberian label kelas data uji pada setiap kelas dengan memberikan nilai keanggotaan seperti halnya teori himpunan *fuzzy*. Algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) diperkenalkan oleh Keller *et al* (1985) dengan mengembangkan K-NN yang digabungkan dengan teori *fuzzy* dalam menyampaikan definisi pemberian label kelas pada data uji yang diprediksi.

Seperti halnya pada teori *fuzzy*, sebuah data mempunyai nilai keanggotaan pada setiap kelas, yang artinya sebuah data bisa dimiliki oleh kelas yang berbeda dengan nilai derajat keanggotaan dalam interval [0,1]. Teori himpunan *fuzzy* menggeneralisasi teori K-NN klasik dengan mendefinisikan nilai keanggotaan sebuah data pada masing-masing kelas. Formula yang digunakan (Liao, 2007):

$$u(x, c_i) = \frac{\sum_{k=1}^K u(x_k, c_i) * d(x, x_k)^{\frac{-2}{(m-1)}}}{\sum_{k=1}^K d(x, x_k)^{\frac{-2}{(m-1)}}} \dots\dots\dots(2.2)$$

Dengan:

$u(x, c_i)$ = Nilai keanggotaan data x ke kelas c_i .

K = Jumlah tetangga yang digunakan.

$u(x_k, c_i)$ = Nilai keanggotaan data tetangga dalam K tetangga pada kelas c_i , nilainya 1 jika data latih x_k milik kelas c_i atau 0 jika bukan milik kelas c_i .

$d(x, x_k)$ = Jarak dari data x ke data x_k dalam K tetangga terdekat.

m = Bobot pangkat yang besarnya >1 .

Nilai keanggotaan suatu data pada kelas sangat dipengaruhi oleh jarak data itu ke tetangga terdekatnya. Semakin dekat ke tetangganya, semakin besar nilai keanggotaan data tersebut. Jarak tersebut diukur dengan N dimensi (fitur) data. Pengukuran jarak (ketidak miripan) dua data. Untuk mendefinisikan jarak antara dua titik yaitu titik pada data training dan titik pada data testing, maka digunakan rumus *Euclidean* :

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots\dots\dots (2.3)$$

Dengan:

x_1 = sampel data

x_2 = data uji

i = variabel data

d = jarak

p = dimensi data

Meskipun FK-NN menggunakan nilai keanggotaan untuk menyatakan keanggotaan data pada setiap kelas, tetapi untuk memberikan keluaran akhir, FK-NN tetap harus memberikan kelas akhir hasil prediksi, untuk keperluan ini, FK-NN memilih kelas dengan nilai keanggotaan terbesar pada data tersebut.

2.8.1 Algoritma Fuzzy K-NN

Berikut ini adalah langkah-langkah yang perlu dilakukan dalam menerapkan metode Fuzzy K-NN dalam pengolahan data (Prasetyo, E.2012). :

1. Normalisasikan data menggunakan nilai terbesar dan terkecil data pada setiap fitur.
2. Cari K tetangga terdekat untuk data uji x menggunakan persamaan (2.2).
3. Hitung nilai keanggotaan $u(x, y_i)$ menggunakan persamaan (2.3) untuk setiap i , dimana $1 \leq i \leq C$.
4. Ambil nilai terbesar $c = u(x, y_i)$ untuk semua $1 \leq i \leq C$, C adalah jumlah kelas.
5. Berikan label kelas c ke data uji x .

2.9 Evaluasi

Menurut Alfiyanto et al (2014) tujuan evaluasi percobaan pada klasifikasi yaitu untuk mengukur keefektifan apakah sistem mengklasifikasi secara benar. Evaluasi biasanya membutuhkan sebuah matriks yang disebut berupa matriks *confusion*. Matriks *counfusion* ditunjukkan pada tabel 2.1.

Tabel 2.1 *Matriks confusion*

Ck	<i>Classifier positive label</i>	<i>Classifier negative label</i>
<i>True positive label</i>	A	B
<i>True negative label</i>	C	D

Dari tabel 2.1 menunjukkan bahwa jika diberikan kategori Ck, parameter A adalah jumlah dokumen yang berhasil dikategorikan oleh sistem ke dalam kategori Ck, parameter B adalah jumlah dokumen yang mempunyai kategori Ck namun sistem tidak mengklasifikasikannya ke dalam kategori Ck, parameter C adalah jumlah dokumen yang bukan kategori Ck namun sistem mengklasifikasikannya ke dalam kategori Ck, dan parameter D adalah jumlah dokumen yang tidak termasuk kategori Ck dan sistem juga tidak mengklasifikasikannya ke dalam kategori Ck.

Recall adalah ukuran keberhasilan sistem dalam mengenali dokumen pada setiap kategori tanpa melihat ketepatan klasifikasi yang dilakukan. *Recall* dihitung menggunakan persamaan 2.4.

$$Recall = \frac{A}{A + B} \dots \dots \dots (2.4)$$

Precision adalah ukuran keberhasilan sistem dalam melakukan ketepatan klasifikasi tanpa melihat seberapa banyak dokumen yang berhasil dikenali. *Precision* dihitung menggunakan persamaan 2.5.

$$Precision = \frac{A}{A + C} \dots \dots \dots (2.5)$$

F-measure mewakili pengaruh relatif antara *precision* dan *recall*, yang dihitung dengan persamaan berikut. *F-measure* dihitung menggunakan persamaan 2.6.

$$Fmeasure = \frac{2 \cdot recall \cdot precision}{(recall + precision)} \dots \dots \dots (2.6)$$

2.10 Penelitian Sebelumnya

Penelitian sebelumnya dilakukan oleh Sugarwanto Atmaja (2016) dengan judul “Sistem Pendukung Keputusan Untuk Deteksi Dini Risiko Penyakit Stroke

Menggunakan *Learning Vector Quantization*". Penelitian yang dilakukan untuk mengklasifikasi pasien yang mengalami risiko penyakit stroke dengan menggunakan atribut meliputi tekanan darah, kadar gula darah, kolesterol total, kolesterol *Low Density Lipoprotein* (LDL), usia, asam urat, jenis kelamin, *Blood Urea Nitrogen* (BUN), dan kreatinin. Nilai parameter yang digunakan pada algoritma LVQ ini meliputi laju pembelajaran 0.04 atau 0.05, fungsi pembelajaran 0.06, jumlah iterasi 7, minimal alfa 0.0001 dan bobot yang digunakan 0.5. Dalam memprediksi status deteksi risiko penyakit stroke pasien dengan nilai akurasi total mencapai 82%.

Penelitian selanjutnya dilakukan oleh Hariyono Hanafi (2015) dengan judul "Penentuan Klasifikasi Status Gizi Orang Dewasa Dengan Menggunakan Metode ANN *Learning Vector Quantization* (LVQ)". Penelitian yang dilakukan untuk mengklasifikasikan status gizi orang dewasa dengan menggunakan data antropometri meliputi data usia, berat badan, tinggi badan, lingkar pinggang dan lingkar pinggul. Nilai parameter yang digunakan pada algoritma LVQ ini meliputi nilai *learning rate* (α) = 0.3, nilai minimal *learning rate* (Mina) = 0.001, nilai pengurangan α adalah 0.6, dan menggunakan 8 neuron dalam melakukan klasifikasi status gizi orang dewasa dengan nilai akurasi mencapai 86%

Penelitian selanjutnya adalah tentang metode FK-NN (*Fuzzy K-Nearest Neighbor*) dalam penelitian yang berjudul "Aplikasi Diagnosa Jenis Baja Berdasarkan Komposisi Kimia Dengan Menggunakan Metode FK-NN (*Fuzzy K-Nearest Neighbor*)", dibuat oleh Muhammad Sholikhuddin (Universitas Muhammadiyah Gresik, 2015). Tujuan dari penelitian tersebut adalah membuat sistem untuk mengetahui jenis baja sesuai dengan komposisi senyawa baja. Atribut-atribut yang digunakan dalam pemilihan baja yaitu senyawa : *Carbon, Silicon, Mangan, Fosfor, Sulfur, Chromium, Molybdenum, Wolfram, Vanadium dan Ferro* *rtc*. Hasil dari penelitian tersebut, Sistem dapat memprediksi data jenis baja. Hasil akurasi dari penelitian tersebut adalah 96,7% yang didapatkan dari 1-NN 4-NN dan 7-NN.