

BAB II

KAJIAN PUSTAKA

Peneliti mengambil sebuah penelitian terdahulu untuk dijadikan dasar penelitian saat ini, yaitu penelitian yang dilakukan oleh Reviangga Dika Satyatama [REV-13]. Dalam penelitian yang dilakukan oleh Reviangga Dika Satyatama [REV-13], penelitian tersebut menerapkan algoritma *Voting Feature Interval-5* (VFI5). Penelitian ini bertujuan untuk menerapkan metode *Fuzzy k-Nearest Neighbor* (Fk-NN) yang dapat diterapkan pada dataset *Indian Liver Patient Dataset* (ILPD). Terdapat dua proses utama yang dilakukan, yang pertama yaitu proses normalisasi nilai atribut pada dataset dan yang kedua yaitu proses pengklasifikasian hasil normalisasi nilai atribut dengan algoritma Fk-NN yang dimana didalamnya terdapat proses yang menghasilkan nilai derajat keanggotaan. Dipilihnya nilai derajat keanggotaan yang terbesar dapat menjadi penentu kelas target yang baru.

2.1. Liver

Liver atau hati adalah organ terbesar dalam tubuh padat dan juga dianggap kelenjar. *Liver* ini terletak di bagian kanan atas perut yang dilindungi oleh tulang rusuk. Penyakit *liver* adalah setiap gangguan fungsi *liver* yang menyebabkan penyakit. *Liver* bertanggung jawab untuk fungsi-fungsi kritis dalam tubuh, hilangnya fungsi-fungsi dapat menyebabkan kerusakan yang signifikan pada tubuh.

Liver adalah satu-satunya organ dalam tubuh yang dapat dengan mudah mengganti sel yang rusak, tetapi jika sel-sel hilang, *liver* tidak mungkin dapat memenuhi kebutuhan tubuh. Kerusakan pada *liver* dapat terjadi dengan berbagai cara, antara lain [SAL-13]:

1. Terjadi peradangan pada sel
2. Aliran empedu terhambat
3. Kolesterol atau *trigliserida* menumpuk
4. Jaringan *liver* rusak oleh bahan kimia dan mineral, atau disusupi oleh sel-sel abnormal

2.1.1. Penyebab Penyakit Liver

Penyakit *liver* dapat disebabkan oleh beberapa hal, antara lain [SAL-13]:

1. Penyalahgunaan alkohol

Alkohol secara langsung merupakan racun bagi sel hati dan dapat menyebabkan peradangan hati yang biasa disebut sebagai hepatitis alkoholik [SAL-13].

2. Sirosis

Sirosis adalah tahap akhir dari penyakit hati. Jaringan parut pada hati dan hilangnya fungsi sel hati dapat menyebabkan *liver* untuk gagal [SAL-13].

3. Jamur beracun

Mengonsumsi jamur tak dikenal yang terkumpul di alam liar dapat mematikan yang menyebabkan timbulnya penyakit *liver* karena jamur yang dikonsumsi merupakan jamur beracun [SAL-13].

4. Infeksi hepatitis

- Hepatitis A adalah infeksi virus yang menyebar terutama melalui rute *fecal-oral* ketika sejumlah kecil dari kotoran yang terinfeksi secara tidak sengaja tertelan.
- Hepatitis B ditularkan oleh paparan cairan tubuh (jarum dari pecandu obat, darah yang terkontaminasi, dan kontak seksual) dan dapat menyebabkan infeksi akut, tetapi juga dapat berkembang menjadi kronis dan dapat menyebabkan peradangan (hepatitis kronis) yang menyebabkan sirosis dan kanker *liver*.
- Hepatitis C menyebabkan hepatitis kronis.
- Hepatitis D adalah virus yang membutuhkan infeksi bersamaan dengan hepatitis B untuk bertahan hidup, dan menyebar melalui paparan cairan tubuh (jarum dari pecandu obat, darah yang terkontaminasi, dan kontak seksual).
- Hepatitis E adalah virus yang menyebar melalui paparan makanan dan air yang terkontaminasi [SAL-13].

5. Virus Lain

Virus lain juga dapat menyebabkan peradangan *liver* atau hepatitis sebagai bagian dari *cluster* gejala. Infeksi virus dengan infeksi *mononucleosis* (virus *Epstein Barr*), *adenovirus*, dan virus *sitomegalo* dapat menyebabkan *liver*. Non-virus

infeksi seperti *toksoplasmosis* dan *Rocky Mountain spotted fever* adalah penyebab kurang umum dari penyakit *liver* [SAL-13].

6. Akumulasi Lemak

NASH atau *non-alkohol steatohepatitis* (juga disebut sebagai "*fatty liver*") menggambarkan akumulasi lemak dalam *liver* yang dapat menyebabkan peradangan *liver* dan penurunan bertahap dalam fungsi *liver* [SAL-13].

7. Hemochromatosis

Hemachromatosis (kelebihan zat besi) adalah gangguan metabolisme yang mengarah ke kadar besi abnormal dalam tubuh. Kelebihan zat besi dapat terakumulasi dalam jaringan dari pankreas jantung, *liver*, dan dapat menyebabkan peradangan, sirosis, kanker *liver*, dan gagal *liver*. *Hemachromatosis* adalah penyakit yang diwariskan [SAL-13].

8. Penyakit Wilson

Penyakit Wilson merupakan penyakit bawaan yang mempengaruhi kemampuan tubuh untuk memetabolisme tembaga. Penyakit Wilson dapat menyebabkan sirosis dan gagal *liver* [SAL-13].

9. Penyakit Gilbert

Penyakit Gilbert adalah kelainan dalam metabolisme *bilirubin* dalam *liver* [SAL-13].

10. Kanker

Kanker primer *liver* timbul dari struktur *liver* dan sel. Contohnya adalah *karsinoma hepatoseluler* dan *cholangiocarcinoma* [SAL-13].

11. Kelainan Aliran Darah

Budd Chiari syndrome adalah penyakit dimana gumpalan darah terbentuk di vena hepatic dan mencegah darah meninggalkan *liver* [SAL-13].

12. Sirosis bilier primer dan *primary sclerosing cholangitis*

Sirosis bilier primer dan *primary sclerosing cholangitis* dapat menyebabkan jaringan parut progresif dari saluran-saluran empedu, menyebabkan saluran empedu menjadi sempit. Sempitnya saluran empedu dapat menyebabkan terhambatnya aliran empedu yang melalui *liver* [SAL-13].

2.1.2. Gejala Penyakit Liver

Gejala penyakit *liver* secara klasik dapat dijelaskan sebagai berikut [SAL-13]:

1. Mual
2. Muntah
3. Sakit pada kuadran kanan atas perut
4. *Jaundice* (perubahan warna kuning pada kulit karena konsentrasi *bilirubin* tinggi dalam aliran darah)
5. Timbulnya gejala kelelahan, lemah, dan penurunan berat badan yang mungkin terjadi.

2.1.3. Diagnosis Penyakit Liver

Untuk mendiagnosa penyakit *liver* dapat dilakukan dengan berbagai cara, antara lain [SAL-13]:

1. Pemeriksaan Fisik

Penyakit *liver* dapat memiliki temuan fisik yang mempengaruhi hampir semua sistem tubuh termasuk jantung, paru-paru, perut, kulit, otak dan fungsi kognitif, dan bagian lain dari sistem saraf. Pemeriksaan fisik sering membutuhkan evaluasi seluruh tubuh.

2. Tes darah

Tes darah sangat membantu dalam menilai peradangan dan fungsi liver. Tes darah untuk fungsi liver meliputi:

- AST dan ALT (bahan kimia *transaminase* dirilis dengan peradangan sel liver);
- GGT dan alkalin fosfatase (bahan kimia yang dikeluarkan oleh sel yang melapisi saluran empedu);
- *Bilirubin*
- Protein dan kadar *albumin*.

3. Pencitraan dapat digunakan untuk memvisualisasikan, tidak hanya liver, tetapi organ terdekat lain yang mungkin sakit. Contoh pencitraan meliputi:

- CT scan.
- MRI (*magnetic resonance imaging*)
- USG (gelombang pencitraan suara, yang sangat membantu dalam menilai kandung empedu dan saluran empedu).

Dalam penelitian ini terdapat beberapa parameter yang digunakan untuk mendiagnosa penyakit *liver*, antara lain:

1. *Usia*

Usia merupakan salah satu faktor yang meningkatkan risiko penyakit *liver*, dimana banyak terjadi pada pasien dengan rentang usia 20 tahun sampai 50 tahun.

2. *Total Bilirubin*

Total bilirubin merupakan jumlah pigmen empedu yang dihasilkan dari pemecahan *heme* dan reduksi *biliverdin*. *Bilirubin* secara normal bersirkulasi di dalam plasma sebagai suatu kompleks dengan *albumin*, diambil oleh sel-sel hati dan dikonjugasikan menjadi *bilirubin diglukuronid* yang merupakan pigmen larut dalam air yang diekskresikan ke dalam empedu. Konsentrasi *bilirubin* yang tinggi dapat menyebabkan ikterus [KUM-98].

3. *Direct Bilirubin*

Direct bilirubin merupakan *bilirubin* yang telah diambil oleh sel-sel hati dan dikonjugasikan membentuk *bilirubin diglukuronid* yang larut dalam air [KUM-98].

4. *Alkaline Phosphotase*

Alkaline phosphotase merupakan *enzim* yang dibuat dalam hati, tulang, dan plasenta dan biasanya hadir dalam konsentrasi tinggi pada darah yang tumbuh didalam empedu dan dalam konsentrasi rendah pada darah. Fosfatase alkali dilepaskan ke dalam darah dalam jumlah yang meningkat selama kerusakan sel-sel hati dan selama aktivitas normal seperti pertumbuhan tulang dan kehamilan. Tingkat abnormal rendah fosfatase alkali hadir dalam kondisi genetik dan hipotiroidisme. Zat ini diukur dalam tes darah rutin [FOS-13].

5. *Alanine Aminotransferase*

Alanine aminotransferase merupakan *enzim* yang secara normal dijumpai pada serum dan jaringan tubuh, terutama pada hati. *Alanine aminotranferase* dilepaskan ke dalam serum sebagai akibat dari kerusakan jaringan, oleh karena itu konsentrasinya dalam serum dapat meningkat pada pasien dengan kerusakan sel hati yang akut [KUM-98].

6. *Aspartate Aminotransferase*

Aspartate aminotransferase merupakan enzim yang biasanya terdapat dalam jaringan tubuh, terutama pada jantung dan hati. *Enzim* itu dilepaskan ke dalam serum sebagai akibat dari cedera jaringan, oleh karena itu konsentrasi dalam serum dapat meningkat pada penyakit seperti infark miokard atau kerusakan akut pada sel-sel hati [KUM-98].

7. *Total Proteins*

Total proteins merupakan jumlah dari setiap kelompok senyawa organik kompleks yang mengandung karbon, hidrogen, oksigen, nitrogen, dan sulfur. Protein, konstituen utama protoplasma semua sel, mempunyai berat molekul yang tinggi dan terdiri atas gabungan asam α -amino dengan ikatan peptida. Dua puluh macam asam amino umumnya ditemukan pada protein, masing-masing protein memiliki rangkaian asam amino yang unik, secara genetik ada batasannya masing-masing serta menentukan bentuk dan fungsi protein yang khas. Protein berfungsi sebagai enzim, unsur struktural, hormon, *imunoglobulin*, dan lain-lain serta terlibat dalam kegiatan transpor oksigen, konsentrasi otot, transpor elektron dan aktivitas lainnya [KUM-98].

8. *Albumin*

Albumin merupakan protein yang larut dalam air dan juga dalam konsentrasi larutan garam yang sedang [KUM-98].

9. *Albumin and Globulin Ratio*

Albumin and globulin ratio merupakan perbandingan *albumin* dengan *globulin*, yang merupakan konstituen utama protein yang ditemukan dalam darah. Rasio yang abnormal terlihat ketika kadar *albumin* atau *globulin* meningkat atau menurun. Rasio abnormal terlihat pada berbagai gangguan, termasuk penyakit ginjal dan hati [ALB-13].

2.2. *Logika Fuzzy*

Logika fuzzy adalah suatu cara yang tepat untuk memetakan ruang input kedalam suatu ruang output [KUS-03]. Konsep ini diperkenalkan dan dipublikasikan pertama kali oleh Lotfi A. Zadeh, seorang profesor dari University of California di Berkeley pada tahun 1965. *Logika fuzzy* menggunakan ungkapan bahasa untuk menggambarkan nilai variabel. *Logika fuzzy* bekerja dengan

menggunakan derajat keanggotaan dari sebuah nilai yang kemudian digunakan untuk menentukan hasil yang ingin dihasilkan berdasarkan atas spesifikasi yang telah ditentukan. Telah disebutkan sebelumnya bahwa logika *fuzzy* memetakan ruang *input* ke ruang *output*. Antara *input* dan *output* ada suatu kotak hitam yang harus memetakan *input* ke *output* yang sesuai. Alasan mengapa orang menggunakan logika *fuzzy*, yaitu [KUS-10]:

1. Konsep logika *fuzzy* mudah dimengerti karena menggunakan konsep matematis yang sederhana dan mudah dimengerti.
2. Logika *fuzzy* sangat fleksibel.
3. Logika *fuzzy* memiliki toleransi terhadap data yang tidak tepat.
4. Logika *fuzzy* mampu memodelkan fungsi-fungsi nonlinier yang sangat kompleks.
5. Logika *fuzzy* dapat membangun dan mengaplikasikan pengalaman-pengalaman para pakar secara langsung tanpa harus melalui proses pelatihan.
6. Logika *fuzzy* dapat bekerjasama dengan teknik-teknik kendali secara konvensional.
7. Logika *fuzzy* didasarkan pada bahasa alami.

2.3. Himpunan *Fuzzy*

Himpunan tegas (*crisp*) A didefinisikan oleh item-item yang ada pada himpunan itu. Pada himpunan tegas (*crisp*), nilai keanggotaan suatu item x dalam suatu himpunan A ($\mu_A(x)$) memiliki dua kemungkinan [KUS-10], yaitu:

1. Satu (1), yang berarti bahwa suatu item menjadi anggota dalam suatu himpunan.
2. Nol (0), yang berarti bahwa suatu item tidak menjadi anggota dalam suatu himpunan.

Himpunan *Fuzzy* didasarkan pada gagasan untuk memperluas jangkauan fungsi karakteristik sedemikian hingga fungsi tersebut akan mencakup bilangan real pada interval $[0,1]$. Nilai keanggotaannya menunjukkan bahwa suatu item dalam semesta pembicaraan tidak hanya berada pada 0 atau 1, namun juga nilai yang terletak diantaranya. Dengan kata lain, nilai kebenaran suatu item tidak hanya benar atau salah. Nilai 0 menunjukkan salah, nilai 1 menunjukkan benar,

dan masih ada nilai-nilai yang terletak antara benar dan salah. Himpunan *fuzzy* memiliki 2 atribut, yaitu [KUS-03]:

- a. Linguistik, yaitu penamaan suatu grup yang mewakili suatu keadaan atau kondisi tertentu dengan menggunakan bahasa alami.
- b. Numeris, yaitu suatu nilai (angka) yang menunjukkan ukuran dari suatu variabel.

Ada beberapa hal yang perlu diketahui dalam memahami sistem *fuzzy*, yaitu:

a) Variabel *Fuzzy*

Variabel *fuzzy* merupakan variabel yang hendak dibahas dalam suatu sistem *fuzzy*.

b) Himpunan *Fuzzy*

Himpunan *fuzzy* merupakan suatu grup yang mewakili suatu kondisi atau keadaan tertentu dalam suatu variabel.

c) Semesta Pembicaraan

Semesta pembicaraan adalah keseluruhan nilai yang diperbolehkan untuk dioperasikan dalam suatu variabel *fuzzy*. Semesta pembicaraan merupakan himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai semesta pembicaraan dapat berupa bilangan positif maupun negatif. Adakalanya nilai semesta pembicaraan ini tidak dibatasi batas atasnya.

d) Domain

Domain himpunan *fuzzy* adalah keseluruhan nilai yang diijinkan dalam semesta pembicaraan dan boleh dioperasikan dalam suatu himpunan *fuzzy*. Seperti halnya semesta pembicaraan, domain merupakan himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai domain dapat berupa bilangan positif maupun negatif.

2.4. Data Mining

2.4.1. Pengertian Data Mining

Data *mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Data *Mining* merupakan suatu proses yang menggunakan teknik statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan

yang terakit dari berbagai database besar [TUR-05]. Data *Mining* merupakan teknologi yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Dengan data *mining* dapat meramalkan *trend* dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh data *mining* melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan. Data *Mining* mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, mencari informasi untuk memprediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka [KHU-07].

Metode data *mining* ini dapat dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, diantaranya yaitu metode deskripsi, estimasi, prediksi, klasifikasi, pengklusteran dan asosiasi [LAR-05].

Pada penelitian ini digunakan metode data *mining* dengan prediksi menggunakan metode gabungan klasifikasi *k-Nearest neighbor* dengan metode *fuzzy*.

Terdapat beberapa fungsionalitas data *mining*, sebagai berikut [HAN-00]:

1. Analisis asosiasi

Menemukan aturan (*rule*) asosiasi yang menunjukkan kondisi nilai atribut yang sering ada bersamaan dalam satu kumpulan data.

2. Klasifikasi

Fungsi pembelajaran yang memetakan (mengklasifikasi) sebuah item data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan.

3. *Clustering*

Melakukan pengelompokan data tanpa berdasarkan kelas data tertentu.

4. *Pendeteksian* perubahan dan deviasi

Berfokus pada penemuan perubahan yang paling signifikan di dalam data dari nilai-nilai yang telah diukur sebelumnya.

2.4.2. Proses Data *Mining*

Prosedur yang umum digunakan untuk permasalahan data *mining* meliputi tahap-tahap sebagai berikut [KAN-03]:

1. Menentukan permasalahan dan merumuskan hipotesis
Pada tahap ini, ditentukan variabel-variabel dan hipotesis awal.
2. Mengumpulkan data
Tahap ini berkaitan dengan bagaimana data dihasilkan dan dikumpulkan.
3. *Preprocessing* data
Dilakukan pembersihan terhadap *outlier*, penanganan *missing value* maupun transformasi data.
4. Memperkirakan model
Pemilihan dan penerapan teknik data *mining* yang sesuai adalah tugas utama dalam tahap ini.
5. Menafsirkan model dan menarik kesimpulan
Pada tahap ini, dilakukan penafsiran model untuk membantu dalam pengambilan keputusan.

Atribut cenderung memiliki nilai dengan rentang yang sangat bervariasi. Misalnya, dalam menentukan jarak antara dua *record*, atribut dengan rentang nilai yang besar, memiliki lebih banyak pengaruh dalam menentukan jarak daripada atribut dengan rentang nilai yang kecil. Oleh karena itu, perlu dilakukan transformasi data berupa normalisasi terhadap nilai untuk membakukan skala pengaruh yang ada pada atribut, terhadap hasil. Ada beberapa teknik normalisasi data seperti normalisasi *min-max* dan Z-score [MOR-09].

Dalam skripsi digunakan normalisasi *min-max*. Normalisasi *min-max* dihitung dengan persamaan 2.1 [MOR-09]. Menurut [JAY-11] normalisasi *min-max* memiliki keunggulan yaitu menjaga relasi pada data. Serta mempunyai fungsi yaitu menyatukan satuan dari berbagai atribut.

$$V' = \frac{V - \min_A}{\max_A - \min_A} \quad (2.1)$$

Dimana,

V' : hasil normalisasi yang nilainya berkisar antara 0 dan 1

V : nilai atribut A yang akan dinormalisasi

\min_A : nilai minimum dari suatu atribut, A

\max_A : nilai maksimum dari suatu atribut, A.

2.5 Klasifikasi

Menurut [KHU-07], klasifikasi adalah suatu fungsionalitas data *mining* yang akan menghasilkan model untuk memprediksi kelas atau kategori dari objek-objek di dalam baris data. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian.

Pada tahap pembelajaran, sebuah algoritma klasifikasi membangun sebuah model klasifikasi dengan cara menganalisis data *training*. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan $y = f(x)$ dimana y adalah kelas hasil prediksi dan x adalah *tuple* yang ingin diprediksi kelasnya.

Selanjutnya, pada tahap pengklasifikasian, model yang telah dihasilkan digunakan untuk melakukan klasifikasi terhadap *unknown* data. Akan tetapi, sebuah model hanya boleh digunakan untuk klasifikasi jika akurasi model tersebut cukup tinggi. Akurasi dapat diketahui dengan cara menguji model tersebut dengan data *test*. Data *test* terdiri dari label kelas yang sudah diketahui, namun data *test* tidak boleh sama dengan data *training* karena menyebabkan pengujian tersebut menunjukkan akurasi yang tinggi, padahal belum tentu demikian.

2.6 Algoritma k-Nearest Neighbor

2.6.1. Definisi k-Nearest Neighbor

Algoritma *k-Nearest Neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *euclidean*.

Beberapa keuntungan dari metode KNN adalah sebagai berikut:

- Sederhana dalam penggunaannya.
- Dapat menangani data training yang mengandung noise.
- Efektif jika data training besar.

Meskipun memiliki beberapa keuntungan metode KNN juga memiliki beberapa kelemahan seperti berikut ini [HAM-08]:

- Computation cost* cukup tinggi karena perlu untuk menghitung jarak setiap data training.
- Membutuhkan memori yang besar.
- Rendahnya tingkat akurasi pada dataset multidimensi.
- Perlu untuk menentukan nilai k parameter, jumlah tetangga terdekat.
- Menggunakan perhitungan jarak, yang belum diketahui pasti jenis jarak yang digunakan.
- Belum diketahui atribut yang lebih baik untuk menghasilkan hasil terbaik.

Kelebihan dari algoritma KNN itu sendiri adalah sederhana dan mudah diimplementasikan. Algoritma ini mencari k *training record* (tetangga) yang memiliki jarak terdekat dari *record* baru, untuk memprediksi kelas dari *record* baru tersebut [YAN-13]. Menurut [MOR-09], KNN juga merupakan algoritma yang sering digunakan untuk klasifikasi pada teknik data *mining*.

2.6.2. Proses K-Nearest Neighbor (KNN)

Menurut Agusta (2007), bahwa prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang dievaluasi dengan k tetangga terdekatnya dalam data pelatihan. Awalnya dicari nilai bobot tiap atribut dari setiap kelas yang ada, kemudian menghitung bobot dengan keanggotaan kelas yang ada pada data latih. Persamaan 2.2 untuk pencarian rata-rata bobot tiap atribut:

$$w = (x_1 + x_2 + \dots + x_n)/n \quad (2.2)$$

Dimana :

w = rata – rata bobot tiap inputan/atribut

x_n = data inputan ke- n tiap atribut

n = jumlah data

Setelah itu dihitung jarak terdekat antara data uji dengan data latih terlebih dahulu. Untuk menghitung jarak digunakan fungsi jarak *euclidean* yang ditunjukkan oleh persamaan 2.3.

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

Keterangan:

w = bobot tiap inputan

x_1 = data latih

x_2 = data uji

i = variabel data

d = jarak

p = dimensi data

Setelah diketahui jarak antar *record*, kemudian diambil sebanyak k tetangga terdekat untuk menentukan label kelas dari *record* baru menggunakan label kelas tetangga.

2.7. Fungsi Kombinasi

Untuk memberikan keputusan klasifikasi bagi *record* baru, dilakukan kombinasi terhadap *record* yang serupa, dengan fungsi kombinasi. Jenis fungsi kombinasi *weighted voting* dilakukan dengan memberikan bobot terhadap beberapa tetangga yang dekat dengan *record* yang baru. Pemberian bobot ini dapat memberikan pengaruh lebih dalam menentukan label kelas. *Weighted* ditunjukkan oleh persamaan 2.4.

$$weight(y) = \frac{1}{distance(y,a)^2} \quad (2.4)$$

Dimana :

y = *neighbor*

a = *new Record*

distance (neighbor, newRecord) merupakan jarak antara *record* yang baru dengan tetangganya. Persamaan 2.4 dapat menghitung bobot dari semua tetangganya, dengan menjumlahkan bobot dari tetangga yang memiliki label kelas yang sama. Label kelas dari *record* yang baru adalah label kelas dari *record* yang jumlah bobot tetangganya paling besar [MOR-09].

2.8. Fuzzy k-Nearest Neighbor

2.8.1. Definisi Fuzzy k-Nearest Neighbor

Fuzzy k-Nearest Neighbor (Fk-NN) merupakan metode klasifikasi yang menggabungkan teknik *fuzzy* dengan *k-Nearest Neighbor classifier*. Algoritma Fk-NN memberikan nilai keanggotaan kelas pada vektor sampel dan bukan

menempatkan vektor pada kelas tertentu. Fk-NN merupakan metode klasifikasi yang digunakan untuk memprediksi data uji menggunakan nilai derajat keanggotaan data uji pada setiap kelas. Kemudian diambil kelas dengan nilai derajat keanggotaan terbesar dari data uji sebagai kelas hasil prediksi. Keuntungannya adalah nilai-nilai keanggotaan vektor seharusnya memberikan tingkat jaminan pada hasil klasifikasi. Sebagai contoh, jika vektor diberikan nilai keanggotaan 0.9 pada kelas pertama dan 0.05 pada dua kelas lainnya peneliti dapat cukup yakin bahwa kelas dengan nilai keanggotaan 0.9 adalah kelas milik vektor tersebut. Di sisi lain, jika sebuah vektor diberi nilai keanggotaan 0.55 pada kelas pertama, 0.44 pada kelas kedua, dan 0.01 pada kelas ketiga, maka peneliti harus ragu-ragu untuk menetapkan vektor berdasarkan hasil ini. Namun, dapat diyakinkan bahwa vektor tersebut bukan milik kelas ketiga. Dalam kasus seperti ini, perlu diperiksa lebih lanjut untuk menentukan klasifikasinya, karena memiliki derajat keanggotaan yang tinggi pada dua kelas yaitu satu dan dua. Pemberian nilai keanggotaan oleh algoritma ini jelas berguna dalam proses klasifikasi.

Sebuah data memiliki nilai keanggotaan pada setiap 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 ditunjukkan oleh persamaan 2.5 [KEL-85].

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} (\|x - x_j\|^{-2/(m-1)})}{\sum_{j=1}^k (\|x - x_j\|^{-2/(m-1)})} \quad (2.5)$$

Dimana u_{ij} adalah nilai keanggotaan fuzzy pada contoh pengujian (x, x_j) yang merupakan satu dari contoh-contoh fuzzy k-nearest neighbor, k merupakan banyaknya nilai ketetanggan terdekat yang diambil, j merupakan variable data untuk keanggotaan data latih, i merupakan variable data untuk keanggotaan data uji, sedangkan m merupakan berat kebalikan yang sebanding dengan jarak Antara x_j dan x . Variabel (m) merupakan penentuan seberapa banyak pemberian bobot pada jarak saat menghitung kontribusi jarak kedekatan pada masing-masing tetangga dengan nilai keanggotaan. Jika nilai m adalah dua, maka jarak kontribusi dari setiap titik tetangga (data latih) dibobotkan oleh nilai kebalikan dari jarak titik tetangga tersebut dengan titik yang sedang diklasifikasikan (data uji). Ketika nilai

m naik, titik-titik tetangga tersebut dibobotkan lebih merata dan efek dari jarak relatif dari titik yang sedang diklasifikasikan akan berkurang. Ketika nilai m mendekati satu, semakin dekat tetangga maka akan dibobotkan lebih besar daripada tetangga yang lebih jauh (semakin besar nilai jarak maka semakin besar bobotnya), yang mana hal ini akan mempengaruhi pengurangan jumlah titik (tetangga) yang berkontribusi terhadap nilai keanggotaan dari titik yang sedang diklasifikasikan. Hasil yang ditampilkan pada jurnal ini, menggunakan nilai $m=2$ tetapi perhatikan bahwa hampir tingkat kesalahan yang diperoleh pada data ini hampir sama dengan menggunakan nilai m yang beragam [KEL85]. Nilai u_{ij} pada $u_i(x)$ terlebih dahulu diproses dengan menggunakan persamaan 2.6 [KEL-85].

$$U_{ij} = \begin{cases} 0.51 + (n_j/K) * 0.49, & \text{jika } j=i \\ (N_j/K) * 0.49, & \text{jika } j \neq i \end{cases} \quad (2.6)$$

Keterangan:

n_j = jumlah anggota kelas j pada suatu dataset K

K = total data latih yang digunakan

j = kelas target (training/tidak training)

2.8.2. Proses Fuzzy K-Nearest Neighbor

Tahapan proses yang dilakukan pada algoritma *Fuzzy K-Nearest Neighbor* ialah:

1. Melakukan normalisasi terhadap nilai-nilai atribut menggunakan normalisasi *min-max* yang ditunjukkan oleh persamaan 2.1.
2. Menghitung jarak antara dua *record* menggunakan *Euclidean distance* yang ditunjukkan oleh persamaan 2.3.
3. Menghitung nilai keanggotaan $u_i(x)$ menggunakan persamaan 2.5 untuk setiap i , dimana $1 \leq i \leq C$, C adalah jumlah kelas.
4. Mengambil nilai terbesar dari proses nomer 3 untuk semua $1 \leq i \leq C$, C adalah jumlah kelas.
5. Memberikan label kelas baru pada proses nomer 4.

2.9. Perhitungan Akurasi

Akurasi merupakan seberapa dekat suatu angka hasil pengukuran terhadap angka sebenarnya (*true value* atau *reference value*). Dalam penelitian ini akurasi diagnosis dihitung dari jumlah diagnosis yang tepat dibagi dengan jumlah data. Tingkat akurasi diperoleh dengan perhitungan sesuai dengan persamaan 2.7 [NUG-06].

$$\text{Akurasi (\%)} = \frac{\sum \text{data uji benar}}{\sum \text{total data uji}} \times 100\% \quad (2.7)$$

Jumlah prediksi benar adalah jumlah *record* data uji yang diprediksi kelasnya menggunakan metode klasifikasi dan hasilnya sama dengan kelas sebenarnya. Sedangkan jumlah total prediksi adalah jumlah keseluruhan *record* yang diprediksi kelasnya (seluruh data uji).

