

### BAB III

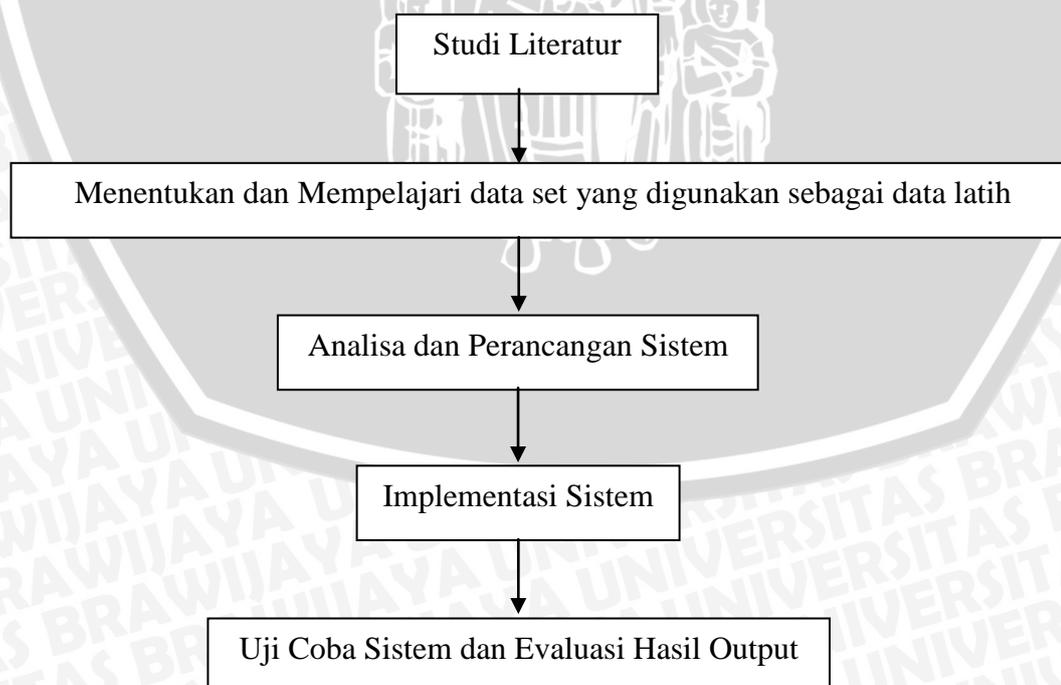
#### METODOLOGI PENELITIAN DAN PERANCANGAN

Dalam bab metodologi dan perancangan ini akan dibahas metode perancangan dan langkah-langkah yang dilakukan dalam penelitian menggunakan metode *Fuzzy k-Nearest Neighbor* untuk diagnose penyakit *liver*.

Penelitian dilakukan dengan tahap-tahap sebagai berikut :

1. Mempelajari Studi literature yang terkait dengan masalah penyakit *liver* dan metode *Fuzzy k-Nearest Neighbor*.
2. Melakukan pengumpulan data – data dari *Indian Liver Patient Dataset (ILPD)*.
3. Menganalisis sistem dari data dan melakukan perancangan sistem yang meliputi pelatihan, dan pengujian.
4. Mengimplementasikan sistem untuk membuat perangkat lunak berdasarkan analisis dan perancangan yang telah dilakukan dengan menggunakan bahasa pemrograman Java.
5. Melakukan uji coba terhadap perangkat lunak.
6. Mengevaluasi uji coba terhadap perangkat lunak.

Langkah-langkah dapat dilihat pada gambar 3.1.



Gambar 3.1 Langkah-langkah penelitian

### 3.1 Studi Literatur

Dalam penelitian ini dibutuhkan studi literatur yang dibutuhkan untuk merealisasikan tujuan dan penyelesaian masalah. Teori – teori mengenai penyakit *liver*, algoritma *fuzzy k-nearest neighbor* digunakan sebagai dasar penelitian yang berasal dari buku – buku, browsing dari internet dan jurnal, baik jurnal nasional maupun jurnal internasional. Kemudian data yang diperoleh diubah sehingga dapat digunakan untuk analisis. Setelah dianalisis maka dapat diimplementasikan ke dalam program.

### 3.2 Data Penelitian

Data yang digunakan dalam penerapan *fuzzy k-nearest neighbor* diambil dari situs <http://archive.ics.uci.edu/ml/datasets.html> dan menggunakan data penyakit *liver* pada tahun 2012. Pada *dataset liver*, atribut yang digunakan adalah usia, jenis kelamin, jumlah *Bilirubin*, *Direct Bilirubin*, *protein total*, *albumin*, rasio A/G, SGPT, SGOT dan Alk Phos. Sedangkan kelas output yaitu:

1. = pasien *liver*.
2. = bukan pasien *liver*.

### 3.3 Analisa dan Perancangan Sistem

#### 3.3.1. Deskripsi Umum Sistem

Secara umum sistem yang dibuat merupakan perangkat lunak yang mengimplementasikan algoritma *Fuzzy k-Nearest Neighbor* (Fk-NN) untuk memprediksi penderita *liver*.

Perangkat lunak ini akan menguji keakuratan hasil klasifikasi *dataset* penderita *liver*. Parameter uji yang berkaitan dengan nilai k (tetangga) dan data latih yang berpengaruh terhadap tingkat akurasi.

#### 3.3.2. Batasan Sistem

Batasan sistem prediksi penderita *liver* ini antara lain:

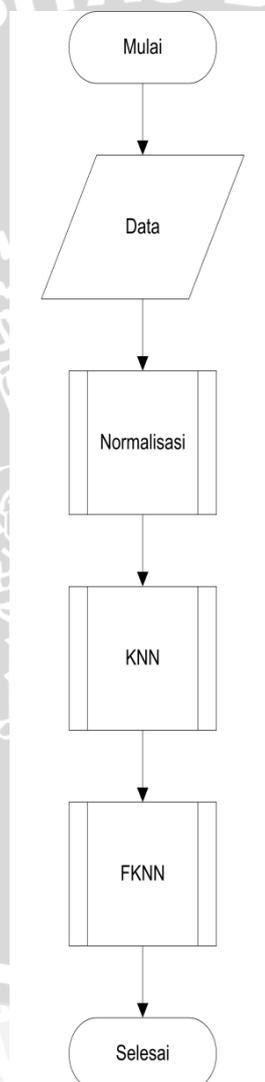
- Input sistem berupa data penderita *liver* selama tahun 2012

### 3.3.3. Perancangan Sistem

Tahap prediksi sistem Fk-NN untuk prediksi penderita *liver*, langkah – langkahnya adalah sebagai berikut:

1. Menginputkan data latih dan data uji penderita *liver* dan  $k$ .
2. Melakukan perhitungan normalisasi untuk nilai setiap atribut.
3. Melakukan perhitungan k-NN.
4. Transformasi output data ke dalam data *fuzzy*.
5. Didapatkanlah hasil prediksi kelas penderita *liver*.

Tahapan proses klasifikasi dapat dilihat pada gambar 3.2 dibawah ini :

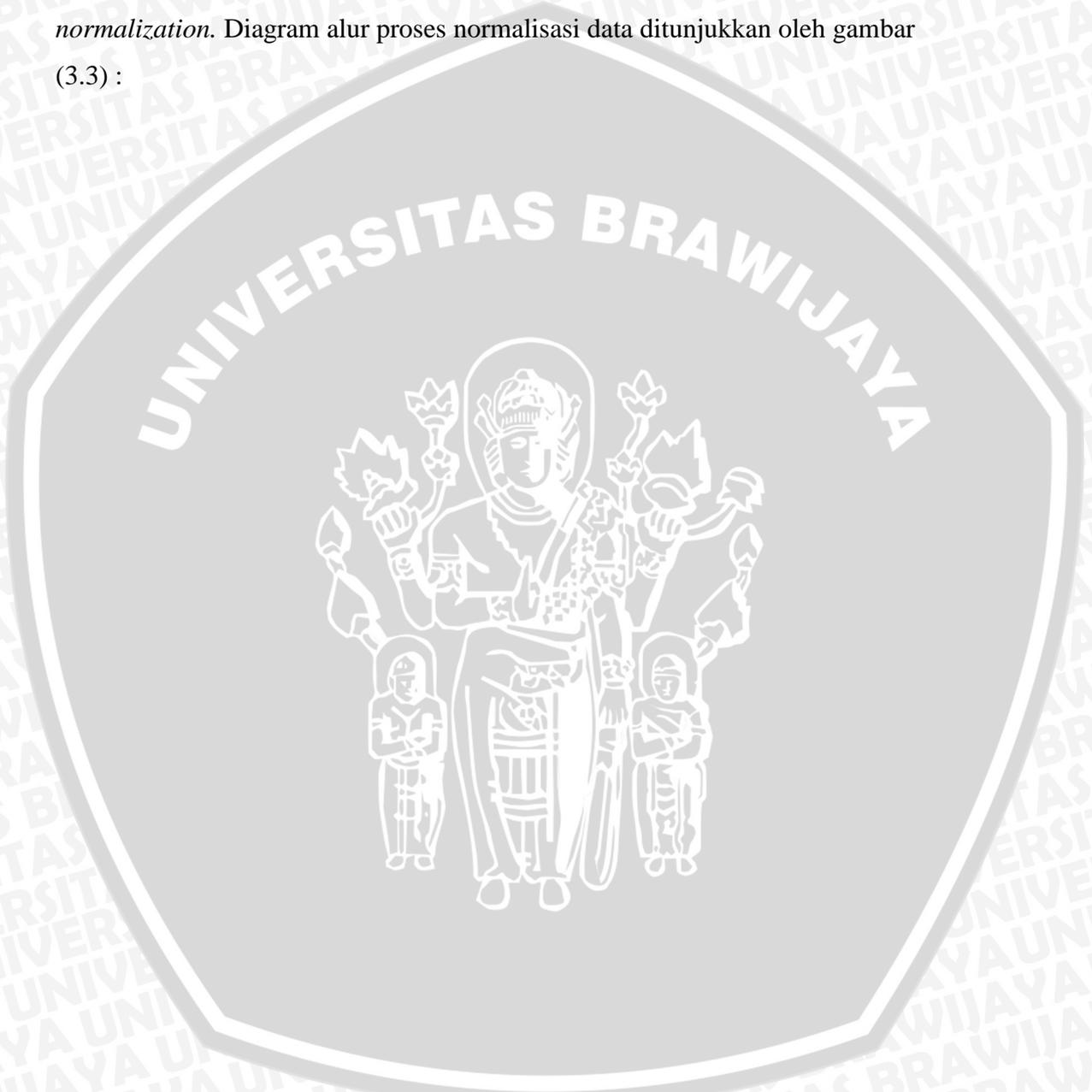


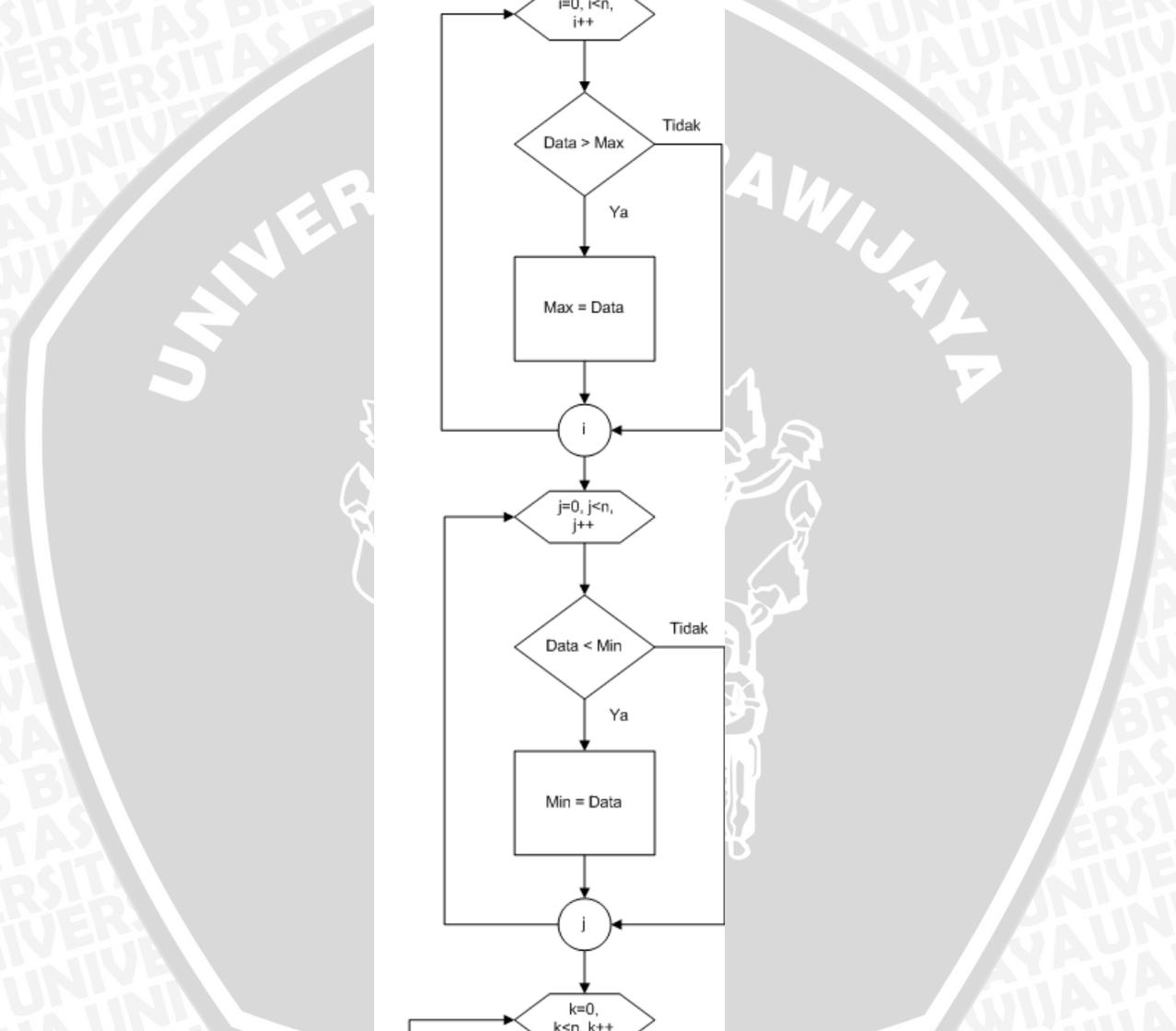
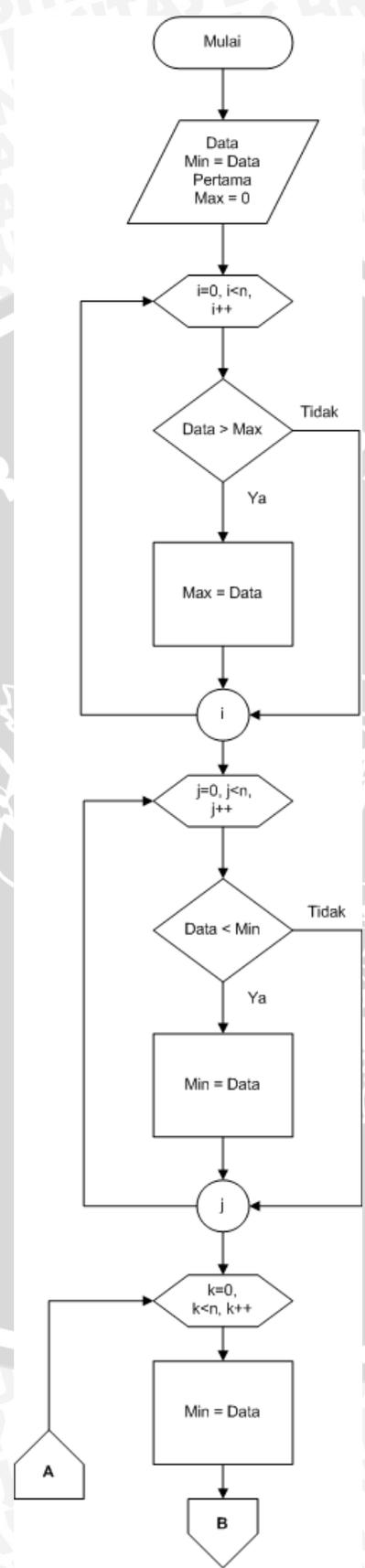
Gambar 3.2 Alur proses klasifikasi

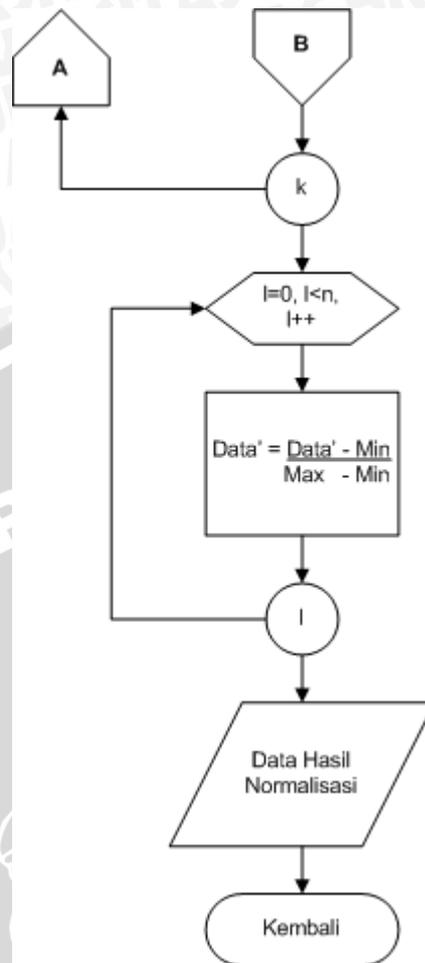
### 3.3.2.1. Normalisasi Data Liver

Proses normalisasi merupakan proses transformasi berupa normalisasi terhadap nilai dari setiap atribut untuk memberlakukan skala pengaruh yang ada pada atribut terhadap hasil. Proses normalisasi yang digunakan adalah *min-max normalization*. Diagram alur proses normalisasi data ditunjukkan oleh gambar

(3.3) :



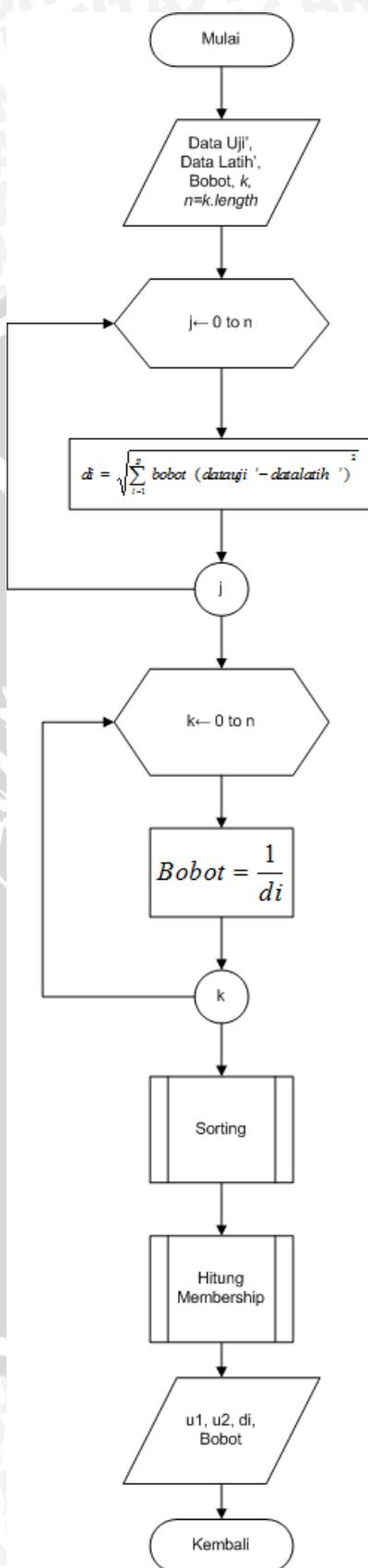




Gambar 3.3 Proses Normalisasi Data

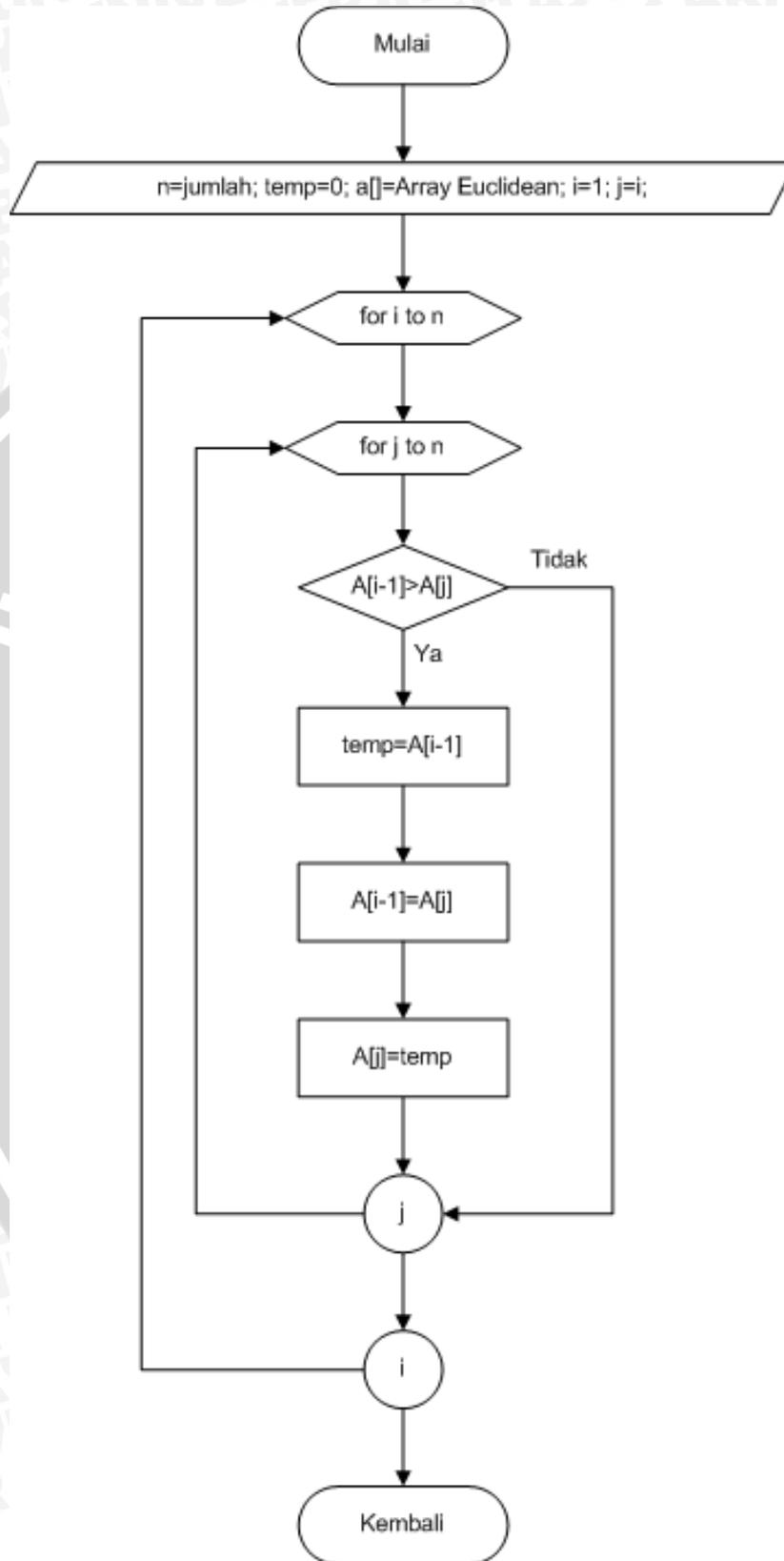
### 3.3.2.2. Proses *K-Nearest Neighbor*

Tahapan ini merupakan proses perhitungan nilai jarak kedekatan tetangga data uji terhadap data latih menggunakan *Euclidean Distance*. Flowchart dari proses *k-nearest neighbor* ditunjukkan oleh gambar (3.4) :



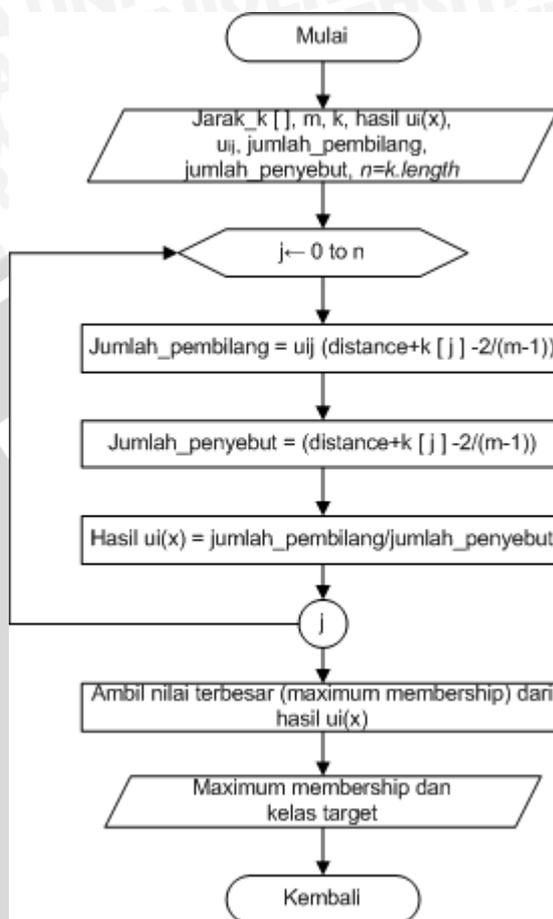
Gambar 3.4 Proses *k*-Nearest Neighbor

Flowchart untuk proses Sorting ditunjukkan oleh gambar (3.5) :



Gambar 3.5 Proses Sorting

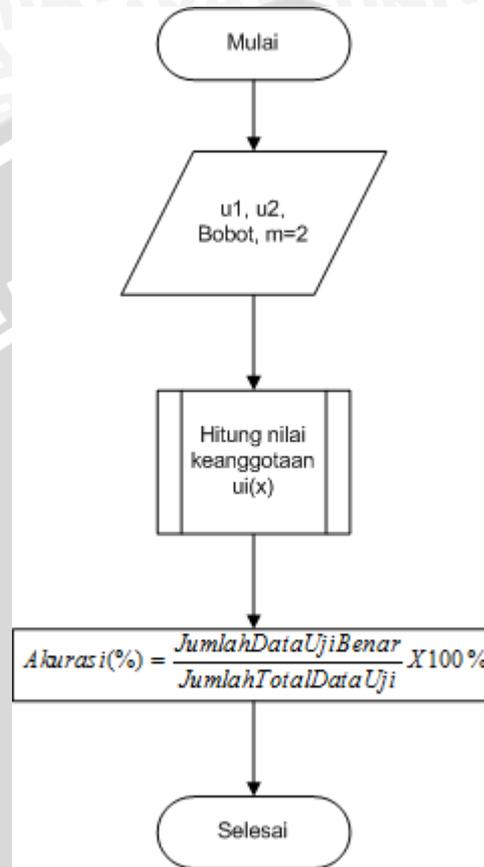
Sedangkan flowchart untuk proses perhitungan membership ditunjukkan oleh gambar (3.6) :



Gambar 3.6 Proses perhitungan membership

**3.3.2.3. Fuzzy k-Nearest Neighbor (Fk-NN)**

Fuzzy k-Nearest Neighbor (Fk-NN) merupakan proses perhitungan nilai membership dengan 2 kelas (penderita liver dan bukan penderita liver). Flowchart untuk proses Fk-NN ditunjukkan oleh gambar (3.7) :



Gambar 3.7 Proses perhitungan fuzzy k-nearest neighbor

**3.4. Contoh Perhitungan Manual**

Pada subbab ini menampilkan perhitungan manual untuk proses diagnosa penyakit liver dan digunakan seluruh atribut dataset tersebut yang berjumlah 10 atribut. Data yang diambil sebanyak 16 record dengan rincian record ke-1 sampai 15 merupakan data latih dan record ke-16 merupakan data uji.

**3.4.1 Data Latih dan Data Uji pada data penderita liver**

Pada contoh perhitungan kali ini digunakan 15 data latih dan 1 data uji. Data latih dan data uji yang dipakai seperti yang ditunjukkan pada tabel (3.1):



Tabel (3.1) Data input latih sistem:

No. Data	Umur	TB	DB	AP	AIA	APA	TP	ALB	AG	Target
1	63	0.9	0.2	194	52	45	6	3.9	1.85	2
2	20	1.1	0.5	128	20	30	3.9	1.9	0.95	2
3	84	0.7	0.2	188	13	21	6	3.2	1.1	2
4	57	1	0.3	187	19	23	5.2	2.9	1.2	2
5	38	2.6	1.2	410	59	57	5.6	3	0.8	2
6	38	2.6	1.2	410	59	57	5.6	3	0.8	2
7	26	0.9	0.2	154	16	12	7	3.5	1	1
8	29	0.9	0.3	202	14	11	6.7	3.6	1.1	1
9	17	0.9	0.3	202	22	19	7.4	4.1	1.2	2
10	55	0.7	0.2	290	53	58	6.8	3.4	1	1
11	57	0.6	0.1	210	51	59	5.9	2.7	0.8	1
12	72	2.7	1.3	260	31	56	7.4	3	0.6	1
13	64	0.9	0.3	310	61	58	7	3.4	0.9	2
14	74	1.1	0.4	214	22	30	8.1	4.1	1	1
15	38	1.8	0.8	342	168	441	7.6	4.4	1.3	1

Keterangan :

Kelas : Pasien Liver = 1; Bukan Pasien Liver = 2

Dari 15 data tersebut, memiliki 2 buah kelas sesuai dengan nilai target output, dimana ke-2 kelas tersebut adalah kelas pasien *liver* dan kelas bukan pasien *liver*.

Tabel (3.2) Data input uji sistem

No	Umur	TB	DB	AP	AIA	APA	TP	ALB	AG	Target
1	25	0.6	0.1	183	91	53	5.5	2.3	0.7	?

### 3.4.2 Melakukan proses normalisasi terhadap nilai atribut

Proses perhitungan normalisasi terhadap nilai atribut ialah dengan menggunakan normalisasi min-max. Sebelum melakukan normalisasi terhadap nilai – nilai atribut, terlebih dahulu dicari nilai minimum dan nilai maksimum dari setiap atribut. Berikut ini adalah contoh perhitungan normalisasi nilai record pertama untuk atribut ke-1 (umur)

$$V = 63 \text{ (nilai record pertama atribut ke-1)}$$

$$\min_A = 17$$

$$\max_A = 84$$

setelah diketahui nilai minimum dan maksimum, kemudian dilakukan perhitungan menggunakan persamaan 2.1

$$v' = \frac{v - \min_a}{\max_a - \min_a}$$

$$= \frac{63 - 17}{84 - 17}$$

$$= 0.687$$

Dari perhitungan ini diperoleh nilai normalisasi untuk record pertama pada atribut ke -1, yaitu 0.687. Proses yang sama juga dilakukan untuk nilai record pada seluruh atribut. Hasil perhitungan normalisasi ditampilkan pada tabel 3.3

Tabel (3.3) Range Data Training

	Umur	TB	DB	AP	AIA	APA	TP	ALB	AG
minimum	17	0.6	0.1	128	13	11	3.9	1.9	0.6
maximum	84	2.7	1.3	410	168	441	8.1	4.4	1.85

Tabel (3.4) Data latih yang telah dinormalisasi

No. Data	Umur	TB	DB	AP	AIA	APA	TP	ALB	AG	Target
1	0.687	0.143	0.083	0.234	0.252	0.079	0.500	0.800	1.000	2
2	0.045	0.238	0.333	0.000	0.045	0.044	0.000	0.000	0.280	2
3	1.000	0.048	0.083	0.213	0.000	0.023	0.500	0.520	0.400	2
4	0.597	0.190	0.167	0.209	0.039	0.028	0.310	0.400	0.480	2
5	0.313	0.952	0.917	1.000	0.297	0.107	0.405	0.440	0.160	2
6	0.313	0.952	0.917	1.000	0.297	0.107	0.405	0.440	0.160	2
7	0.134	0.143	0.083	0.092	0.019	0.002	0.738	0.640	0.320	1
8	0.179	0.143	0.167	0.262	0.006	0.000	0.667	0.680	0.400	1
9	0.000	0.143	0.167	0.262	0.058	0.019	0.833	0.880	0.480	2
10	0.567	0.048	0.083	0.574	0.258	0.109	0.690	0.600	0.320	1
11	0.597	0.000	0.000	0.291	0.245	0.112	0.476	0.320	0.160	1
12	0.821	1.000	1.000	0.468	0.116	0.105	0.833	0.440	0.000	1
13	0.701	0.143	0.167	0.645	0.310	0.109	0.738	0.600	0.240	2
14	0.851	0.238	0.250	0.305	0.058	0.044	1.000	0.880	0.320	1
15	0.313	0.571	0.583	0.759	1.000	1.000	0.881	1.000	0.560	1

Tabel (3.5) Data uji yang telah dinormalisasi

No.	Umur	TB	DB	AP	AIA	APA	TP	ALB	AG
1	0.119	0.000	0.000	0.195	0.503	0.098	0.381	0.160	0.080

Tabel (3.6) Bobot setiap atribut data latih

kelas	Umur	TB	DB	AP	AIA	APA	TP	ALB	AG
1	0.495	0.306	0.310	0.393	0.243	0.196	0.755	0.651	0.297
2	0.457	0.351	0.354	0.445	0.162	0.065	0.461	0.510	0.400

### 3.4.3 Menghitung jarak antara record baru pada data uji dengan tiap record pada data latih

Dari data uji dengan data pelatihan, dicari terlebih dahulu jarak terdekat “Euclidean distance”, dimana rumus untuk Euclidean distance dapat dilihat pada persamaan (2.3). Berikut ini adalah contoh, perhitungan nilai jarak terdekat dari *record* pertama data latih dengan record uji.

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

$$\text{Jarak} = \sqrt{\frac{\begin{aligned} &(0.457*((0.119-0.687)^2))+(0.351*((0-0.143)^2)) \\ &+(0.354*(0-0.083)^2)+(0.445*(0.195-0.234)^2) \\ &+(0.162*((0.503-0.252)^2))+(0.065*((0.098-0.079)^2)) \\ &+(0.461*((0.381-0.500)^2))+(0.510*((0.160-0.800)^2)) \\ &+(0.400*(0.080-1)^2) \end{aligned}}{(0.457+0.351+0.354+0.445+0.162+0.065+0.461+0.510+0.400)}} \\ = 0.699$$

Selanjutnya dihitung lagi jarak antara record uji dengan record data latih yang lain. Hasil perhitungan jarak ditampilkan pada tabel 3.7

Tabel (3.7) Hasil perhitungan jarak

Data Ke-	Nilai Target	Target
1	0.699	2
2	0.445	2
3	0.696	2
4	0.463	2
5	0.985	2
6	0.985	2
7	0.568	1
8	0.571	1
9	0.661	2
10	0.614	1
11	0.396	1
12	1.063	1
13	0.655	2
14	0.993	1
15	1.112	1

Selanjutnya dari data yang telah ditemukan nilai Euclidean distance dilakukan pencarian nilai *weight* (*w*) dengan menggunakan persamaan 2.4. Berikut ini

adalah contoh perhitungan nilai *weight* dari hasil *record* pertama perhitungan Euclidean distance.

$$\begin{aligned} \text{Weight} &= \frac{1}{0.699} \\ &= 2.046 \end{aligned}$$

Untuk hasil perhitungan nilai *weight* dari hasil Euclidean Distance untuk seluruh *record* dapat dilihat pada tabel 3.8

Tabel (3.8) nilai weighted

Data ke-	Nilai	Target
1	2.046	2
2	5.052	2
3	2.063	2
4	4.675	2
5	1.031	2
6	1.031	2
7	3.101	1
8	3.063	1
9	2.289	2
10	2.657	1
11	6.380	1
12	0.884	1
13	2.331	2
14	1.014	1
15	0.809	1

Setelah hasil perhitungan jarak didapatkan maka pada data setelah mengalami proses *weighted* data tersebut diurutkan sesuai dengan nilai yang paling besar.

Jika nilai di ambil dari Euclidean distance yang di ambil nilai yang terkecil, karena *weighted* merupakan nilai invers dari Euclidean distance, maka yang di ambil adalah nilai terbesar. Hasil perhitungan jarak setelah diurutkan dari yang terbesar ditunjukkan pada tabel 3.9

Tabel (3.9) Data setelah diurutkan dari yang terbesar

data ke-	Nilai	Target	Urutan Terbesar
11	6.380	1	1
2	5.052	2	2
4	4.675	2	3
7	3.101	1	4
8	3.063	1	5
10	2.657	1	6

13	2.331	2	7
9	2.289	2	8
3	2.063	2	9
1	2.046	2	10
5	1.031	2	11
6	1.031	2	12
14	1.014	1	13
12	0.884	1	14
15	0.809	1	15

**3.4.4 Menentukan k record terdekat**

Berdasarkan hasil perhitungan *weighted* pada tabel 3.8, kemudian dilakukan pengurutan terhadap *weight record* yang ditunjukkan pada tabel 3.9. kemudian diambil k record dengan berat terbesar. Apabila ditentukan k = 3, maka record yang terpilih adalah record ke-10, 13, dan 8.

Tabel 3.10 seleksi 3 data dengan nilai terbesar

data ke-	Nilai	Target	Urutan Terbesar
11	6.380	1	1
2	5.052	2	2
4	4.675	2	3

**3.4.5 Menentukan maximum membership dan kelas target**

Proses menentukan maximum membership dan kelas target dengan mencari nilai membership untuk tiap kelas j dengan menggunakan persamaan 2.5. K=15 (jumlah data latih),  $u_1$ =pasien liver,  $u_2$ =bukan pasien liver,  $n_1=7$ ,  $n_2=8$

$$\begin{aligned}
 u_{1(1)} &= 0.51 + (7/15) * 0.49 \\
 &= 0.51 + 0.46 * 0.49 \\
 &= 0.51 + 0.229 \\
 &= 0.739
 \end{aligned}$$

$$\begin{aligned}
 u_{1(2)} &= (7/15) * 0.49 \\
 &= 0.46 * 0.49 \\
 &= 0.229
 \end{aligned}$$

$$\begin{aligned}
 u_{2(2)} &= 0.51 + (8/15) * 0.49 \\
 &= 0.51 + 0.53 * 0.49 \\
 &= 0.51 + 0.26 \\
 &= 0.771
 \end{aligned}$$

$$u_{2(1)} = (8/15) * 0.49$$



$$= 0.53 * 0.49$$

$$= 0.261$$

Setelah didapatkan membership untuk tiap kelas  $j$  dilanjutkan dengan mencari nilai keanggotaan sebuah data pada masing – masing kelas dengan menggunakan persamaan 2.5

$$u_1 = \frac{\left(0.739 * \left(6.380^{-\frac{2}{2-1}}\right)\right) + \left(0.229 * \left(5.052^{-\frac{2}{2-1}}\right)\right) + \left(0.229 * \left(4.675^{-\frac{2}{2-1}}\right)\right)}{\left(6.380^{-\frac{2}{2-1}}\right) + \left(5.052^{-\frac{2}{2-1}}\right) + \left(4.675^{-\frac{2}{2-1}}\right)}$$

$$= \frac{0.038}{0.110}$$

$$= 0.343$$

$$u_2 = \frac{\left(0.261 * \left(6.380^{-\frac{2}{2-1}}\right)\right) + \left(0.771 * \left(5.052^{-\frac{2}{2-1}}\right)\right) + \left(0.771 * \left(4.675^{-\frac{2}{2-1}}\right)\right)}{\left(6.380^{-\frac{2}{2-1}}\right) + \left(5.052^{-\frac{2}{2-1}}\right) + \left(4.675^{-\frac{2}{2-1}}\right)}$$

$$= \frac{0.072}{0.110}$$

$$= 0.657$$

Berdasarkan hasil perhitungan nilai keanggotaan didapat dua nilai keanggotaan, untuk menentukan kelas target maka dipilih nilai keanggotaan terbesar yaitu 0.657 sehingga kelas targetnya yaitu 2 (bukan penderita liver). Jadi, status diagnosa penderita *liver* yang didapat ialah penderita tersebut tidak menderita penyakit *liver*.

Setelah melakukan pengujian data uji langkah yang sama dilakukan terhadap data uji yang lain sehingga didapat akurasi sistem yang dihitung berdasarkan rumus (2.7)

Tabel 3.11 Perhitungan Akurasi Sistem

data uji ke-	Kelas Data	Kelas Hasil Perhitungan	Hasil Prediksi
1	2	2	Benar
2	1	2	Salah
3	2	2	Benar

$$\text{Akurasi (\%)} = \frac{2}{3} \times 100\%$$

$$= 66,66667 \%$$

### 3.5 Perancangan Antarmuka

Antarmuka (*interface*) untuk menampilkan data yang digunakan sistem, terdiri dari 3 bagian, yaitu Tampil Data Latih untuk memanggil dan melatih data latih, Tampil Data Uji untuk memanggil dan menguji data uji, dan FKNN untuk menampilkan nilai k yang dimasukkan beserta hasil akurasi. Perancangan antarmuka sistem akan ditunjukkan pada gambar berikut ini.

Tampil Data Latih Tampil Data Uji FKNN

Browse

Data Latih

Title 1	Title 2	Title 3	Title 4

Data Latih yang telah dinormalisasi

Title 1	Title 2	Title 3	Title 4

Gambar 3.8 Tampil Data Latih

1. Tombol Browse digunakan untuk mencari data latih yang ingin digunakan.
2. Tabel Data Latih digunakan untuk menampilkan data latih yang dipilih saat menekan tombol Browse.
3. Tabel Data Latih yang telah dinormalisasi digunakan untuk menampilkan hasil normalisasi dari data latih yang dipilih.

The screenshot shows a software window titled 'Tampil Data Uji' with three tabs: 'Tampil Data Latih', 'Tampil Data Uji', and 'FKNN'. The 'Tampil Data Uji' tab is active. It contains a 'Browse' button and two data tables. The first table, labeled 'Data Uji', has four columns: 'Title 1', 'Title 2', 'Title 3', and 'Title 4', with three empty rows below. The second table, labeled 'Data Uji yang telah dinormalisasi', also has four columns: 'Title 1', 'Title 2', 'Title 3', and 'Title 4', with three empty rows below.

Gambar 3.9 Tampil Data Uji

1. Tombol Browse digunakan untuk mencari data uji yang ingin digunakan.
2. Tabel Data Uji digunakan untuk menampilkan data uji yang dipilih saat menekan tombol Browse.
3. Tabel Data Uji yang telah dinormalisasi digunakan untuk menampilkan hasil normalisasi dari data uji yang dipilih.

The screenshot shows a software window titled 'FKNN' with three tabs: 'Tampil Data Latih', 'Tampil Data Uji', and 'FKNN'. The 'FKNN' tab is active. It contains two input fields: 'Nilai k' with an adjacent 'OK' button, and 'AKURASI =' followed by an empty text box.

Gambar 3.10 FKNN

1. *Textbox* untuk menampilkan nilai k yang dimasukkan.
2. Tombol OK digunakan untuk melakukan proses Fuzzy KNN.

3. *Textfield* untuk menampilkan hasil akurasi hasil pengujian dari proses Fuzzy KNN.

### 3.6 Perancangan Uji Coba

Setelah sistem selesai dibuat, langkah selanjutnya adalah melakukan pengujian terhadap sistem tersebut. Pengujian dilakukan untuk mengetahui tingkat akurasi dari hasil klasifikasi pada *Indian Liver Patient Dataset* (ILPD) dengan menggunakan metode *Fuzzy k-Nearest Neighbor*.

Terdapat 1 macam pengujian yang dilakukan dalam penelitian ini, yaitu: pengujian untuk mengetahui pengaruh nilai k (jumlah tetangga terdekat) dan data latih terhadap tingkat akurasi klasifikasi.

### 3.7 Uji pengaruh nilai k dan data latih terhadap tingkat akurasi klasifikasi

Uji pengaruh nilai k dan jumlah data latih terhadap tingkat akurasi klasifikasi dilakukan pada sejumlah data uji yang sama yaitu 50 data uji dengan beberapa data latih yang berbeda-beda. Dengan beberapa data latih yang berbeda-beda, tentunya akan mempengaruhi keakuratan hasil keputusan yang berpengaruh terhadap penentuan kelas. Dalam pengujian ini ditentukan nilai  $k = 2 \dots n$ . Dari data latih diambil sejumlah record, meliputi 80, 110, 140, 170, 200, dan 230 record.

Dari pengujian ini, diperoleh tingkat akurasi terhadap nilai k dan jumlah data latih. Tabel 3.12 menampilkan rancangan tabel yang akan digunakan untuk mencatat hasil dari pengujian ini.

Tabel 3. 12 Uji Pengaruh nilai k dan data latih terhadap tingkat akurasi

K	Jumlah Data Latih	Akurasi (%)
2	80	
...	...	
n	n	

Keterangan:

k : nilai k yang akan diuji.

Jumlah Data Latih : jumlah data yang digunakan sebagai data latih.

Akurasi : tingkat akurasi dalam persen, yang dihitung dengan persamaan 2.7.