

BAB V

PENGUJIAN DAN ANALISIS

Bab ini berisi tentang pengujian dan analisis hasil pengujian sistem dalam penentuan tingkat resiko penyakit lemak darah dengan menggunakan metode *Modified k-Nearest Neighbor* (MKNN).

5.1 Skenario Uji Coba

Skenario uji coba dalam penentuan tingkat resiko penyakit lemak darah ini, digunakan 200 *dataset* penyakit lemak darah. Dari *dataset* penyakit lemak darah tersebut digunakan 4 parameter untuk menentukan tingkat resiko penyakit, yaitu kadar kolesterol total, kolesterol HDL, kolesterol LDL dan trigliserida.

Untuk skenario uji coba yang dilakukan, yaitu dengan jumlah data latih dan data uji yang telah ditentukan, akan dibagi secara acak dengan persentase datanya sebesar 140 data latih dengan 60 data uji, 100 data latih dengan 100 data uji dan 60 data latih dengan 140 data uji. Data latih sebesar 140 diasumsikan untuk mewakili jumlah data uji kecil dan jumlah data latih yang lebih besar. Data latih sebesar 100 diasumsikan untuk jumlah data uji dan data latih yang sama. Sedangkan data latih sebesar 60 diasumsikan untuk mewakili jumlah data uji yang lebih besar daripada jumlah data latih. Pengujian sistem ini dilakukan dengan nilai k yang berbeda, yaitu $k=1$ hingga $k=10$. Setiap nilai k akan dilakukan uji coba sebanyak 5 kali, dan didapatkan nilai akurasi rata-rata.

5.1.1 Pengujian Pada Jumlah 140 Data Latih

Pada pengujian data penyakit lemak darah dengan jumlah latih sebesar 140 dihasilkan rata-rata nilai akurasi yang berbeda-beda. Untuk pengujian data uji sebesar 60 data dihasilkan rata-rata nilai akurasi tertinggi sebesar 86,00% pada nilai $k=2$ dan dihasilkan pula rata-rata nilai akurasi terendah sebesar 80,03% pada nilai $k=9$. Besar jumlah total akurasi rata-rata hasil pengujian pada jumlah 140 data latih sebesar 81,98%. Hasil uji coba pada jumlah data latih sebesar 140 ditunjukkan pada Tabel 5.1

Tabel 5.1. Hasil Uji Coba pada Jumlah Data Latih 140

Jumlah Latih	Nilai Tetangga (K)	Nilai Akurasi Data Uji (%)
		60
140	1	84,56%
	2	86,00%
	3	83,84%
	4	80,45%
	5	82,25%
	6	80,32%
	7	80,03%
	8	81,75%
	9	80,03%
	10	81,81%
Akurasi Rata-Rata (%)		81,98%

5.1.2 Pengujian Pada Jumlah 100 Data Latih

Pada pengujian data penyakit lemak darah dengan jumlah latih sebesar 100 dihasilkan rata-rata nilai akurasi yang berbeda-beda. Untuk pengujian data uji sebesar 100 data dihasilkan rata-rata nilai akurasi tertinggi sebesar 77,17% pada nilai $k=2$ serta dihasilkan pula rata-rata nilai akurasi terendah sebesar 71,53% pada nilai $k=9$. Besar jumlah total akurasi rata-rata hasil pengujian pada jumlah 100 data latih sebesar 73,50%. Hasil uji coba pada jumlah data latih sebesar 100 ditunjukkan pada tabel 5.2.

Tabel 5.2. Hasil Uji Coba pada Jumlah Data Latih 100

Jumlah Latih	Nilai Tetangga (K)	Nilai Akurasi Data Uji (%)
		100
100	1	75,56%
	2	77,17%
	3	74,36%
	4	73,56%
	5	73,09%
	6	72,31%
	7	72,27%
	8	72,96%
	9	71,53%
	10	72,18%
Akurasi Rata-Rata (%)		73,50%

5.1.3 Pengujian Pada Jumlah 60 Data Latih

Pada pengujian data penyakit lemak darah dengan jumlah latih sebesar 40 dihasilkan rata-rata nilai akurasi yang berbeda-beda. Untuk pengujian data uji sebesar 60 data dihasilkan rata-rata nilai akurasi tertinggi sebesar 72,67% pada nilai $k=2$ dan dihasilkan pula rata-rata nilai akurasi terendah sebesar 64,34% pada nilai $k=9$. Besar jumlah total akurasi rata-rata hasil pengujian pada jumlah 40 data latih sebesar 68,82%. Hasil uji coba pada jumlah data latih sebesar 40 ditunjukkan pada Tabel 5.3.

Tabel 5.3. Hasil Uji Coba pada Jumlah Data Latih 40

Jumlah Latih	Nilai Tetangga (K)	Nilai Akurasi Data Uji (%)
		140
60	1	72,22%
	2	72,67%
	3	71,70%
	4	71,56%
	5	70,00%
	6	70,67%
	7	66,06%
	8	64,54%
	9	64,34%
	10	64,44%
Akurasi Rata-Rata (%)		68,82%

5.1.4 Pengujian Terhadap Data Latih Dengan Nilai K Terbaik

Pada pengujian ini, akan digunakan nilai k terbaik yang didapatkan dari pengujian sebelumnya, yaitu nilai $k=2$. Untuk jumlah data latih yang digunakan adalah 30%, 50% dan 70% dari jumlah *dataset*. Pengujian pada jumlah data latih sebesar 30% data dihasilkan rata-rata nilai akurasi tertinggi sebesar 73,55%, pada jumlah data latih sebesar 50% data dihasilkan rata-rata nilai akurasi tertinggi sebesar 78,17% dan pada pengujian data latih sebesar 70% adalah 85,81%. Hasil uji coba data latih dengan nilai k terbaik ditunjukkan pada Tabel 5.4.

Tabel 5.4. Hasil Uji Coba Data Latih Dengan Nilai K Terbaik

Jumlah <i>Dataset</i>	Jumlah Data Latih	Nilai K	Akurasi (%)
200	60	2	73,55%
	100		78,17%
	140		85,81%

5.1.5 Pengujian Perbandingan *Dataset* dengan Kelas yang Seimbang (*Balanced Class*) dan Tidak Seimbang (*Imbalanced Class*).

Pengujian dengan jumlah kelas *dataset* yang seimbang yang terdiri dari 50 kelas kategori normal, 50 kelas kategori waspada serta 50 kelas kategori tingkat resiko tinggi. Pada pengujian ini digunakan jumlah *dataset* yang berbeda, yaitu sebesar 50, 75, dan 100 dengan persentase jumlah data latih sebesar 50%.

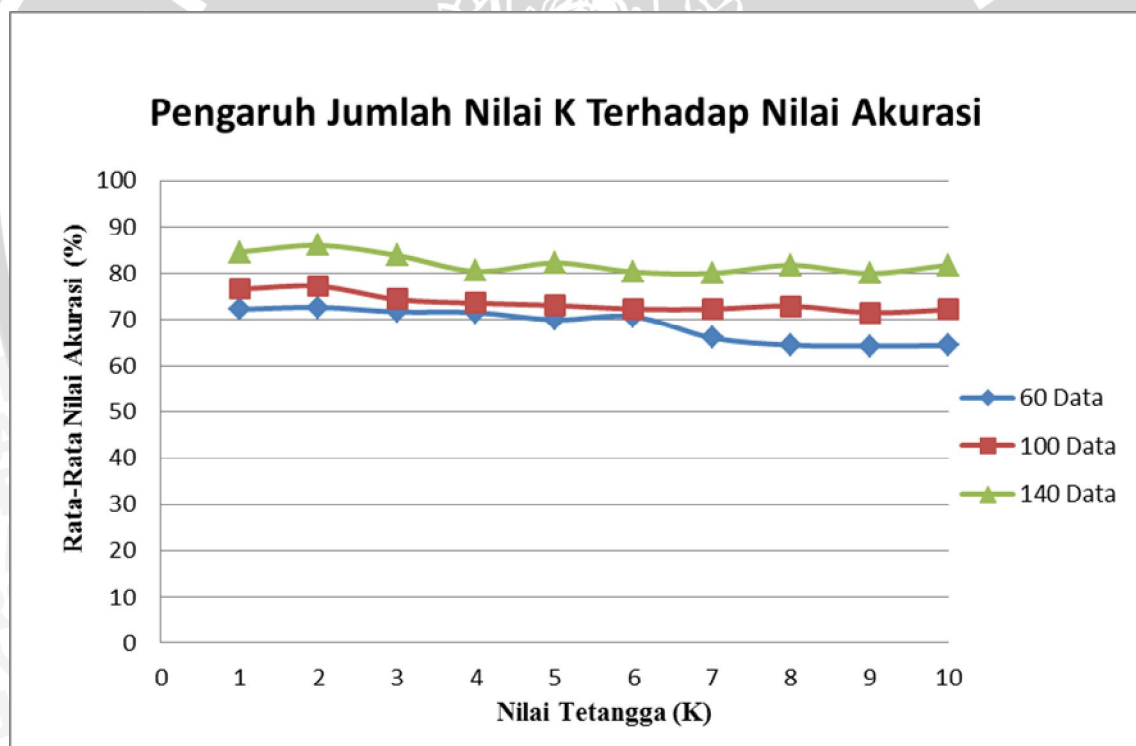
Pada pengujian dengan data yang tidak seimbang (*imbalanced class*) untuk jumlah *dataset* sebesar 50 dihasilkan akurasi rata-rata sebesar 65,98%, jumlah *dataset* 75 dihasilkan rata-rata akurasi sebesar 63,35%, dan untuk jumlah *dataset* 100 dihasilkan rata-rata akurasi sebesar 60,23%. Sedangkan untuk *dataset* yang seimbang (*balanced class*) untuk jumlah *dataset* sebesar 50 dihasilkan akurasi rata-rata sebesar 74,44%, jumlah *dataset* 75 dihasilkan rata-rata akurasi sebesar 78,87%, dan untuk jumlah *dataset* 100 dihasilkan rata-rata akurasi sebesar 82,56%. Hasil uji coba pada *dataset* dengan kelas yang seimbang dan tidak seimbang ditunjukkan pada Tabel 5.5. Tabel 5.5 Hasil Uji Coba Pada *Dataset* dengan Kelas yang Seimbang dan Tidak Seimbang.

Jenis Data	Jumlah Dataset	Jumlah Data Latih	Nilai Akurasi Rata-Rata (%)
Tidak Seimbang	50	50%	65,98%
	75		63,35%
	100		60,23%
Seimbang	50	50%	74,44%
	75		78,87%
	100		82,56%

5.2 Analisis Hasil Uji Coba

5.2.1 Analisis Pengaruh Jumlah Nilai k Terhadap Akurasi Sistem

Berdasarkan hasil uji coba dengan menggunakan metode MKNN, pada masing-masing jumlah uji yang telah diuji dengan perubahan nilai k memberikan pengaruh pula terhadap nilai akurasi yang dihasilkan. Akan tetapi rata-rata nilai akurasi cenderung semakin menurun sejalan dengan adanya penambahan nilai k . Hal ini dikarenakan adanya kelas yang mendominasi *dataset* tersebut, sehingga untuk data yang diambil pada perhitungan memiliki kelas yang sama. Pada uji coba dengan 200 *dataset* tersebut cenderung pada $k=2$, karena nilai k yang kecil akan mengurangi *noise*. Grafik pengaruh nilai k terhadap akurasi ditunjukkan pada Gambar 5.1.

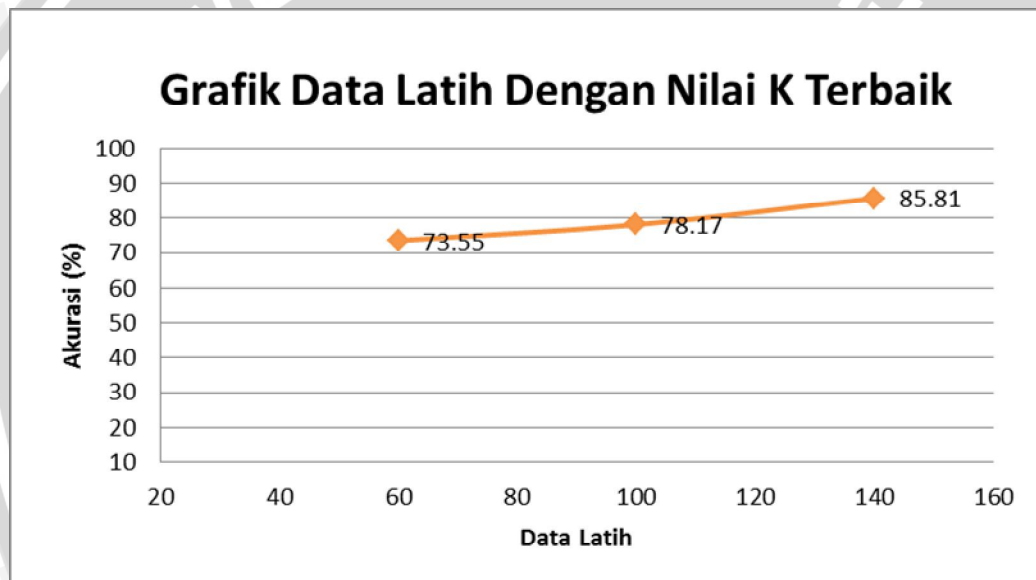


Gambar 5.1 Grafik Pengaruh Jumlah Nilai K terhadap Nilai Akurasi

5.2.2 Analisis Pengujian pada Jumlah Data Latih Dengan Nilai K Terbaik

Berdasarkan hasil uji coba yang telah dilakukan, terlihat bahwa jumlah data latih juga berpengaruh pada besarnya nilai akurasi yang dihasilkan. Pengujian

pada jumlah data latih sebesar 30% data dihasilkan rata-rata nilai akurasi tertinggi sebesar 72,60%, pada jumlah data latih sebesar 50% data dihasilkan rata-rata nilai akurasi tertinggi sebesar 77,17% dan pada pengujian data latih sebesar 70% adalah 86%. Peningkatan jumlah data latih turut disertai dengan meningkatnya nilai akurasi, sehingga semakin banyaknya data latih maka akan berpengaruh pada besarnya nilai akurasi yang dihasilkan. Selain itu, jika kecilnya data latih dan semakin banyaknya jumlah data uji, kemungkinan semakin sedikitnya jarak yang mendekati kelas data prediksi. Grafik Jumlah Data latih dengan nilai K terbaik ditunjukkan pada Gambar 5.2.

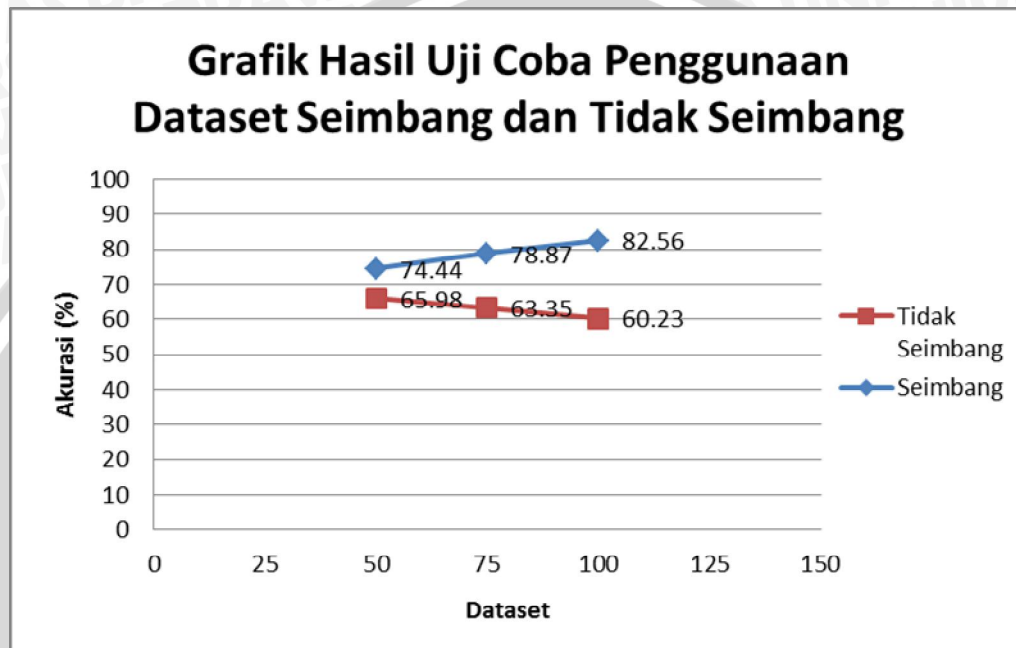


Gambar 5.2 Grafik Pengaruh Jumlah Data Latih Menggunakan Nilai K Terbaik

5.2.3 Analisis Pengujian Perbandingan *Dataset* dengan Kelas yang Seimbang (*Balanced Class*) dan Tidak Seimbang (*Imbalanced Class*)

Pada pengujian untuk jenis data dengan kelas yang tidak seimbang (*imbalanced class*) dengan peningkatan jumlah *dataset* yang digunakan, hal tersebut turut disertai dengan penurunan nilai akurasi. Hal ini sebaliknya ditunjukkan pula pada jenis *dataset* dengan kelas yang seimbang pada peningkatan jumlah *dataset* yang digunakan turut disertai dengan peningkatan nilai akurasi. Hal ini disebabkan karena *dataset* yang tidak seimbang menimbulkan banyak data yang tidak relevan

(noise) dalam penentuan keputusan. Pada *dataset* yang tidak seimbang keputusan lebih cenderung mengacu pada kelas yang mendominasi dalam *dataset*. Grafik perbandingan akurasi pada *dataset* dengan kelas yang seimbang dan kelas yang tidak seimbang ditunjukkan pada Gambar 5.3.



Gambar 5.3 Grafik Hasil Uji Coba Dengan Jumlah Dataset Seimbang dan Tidak Seimbang