

BAB V

PENGUJIAN DAN ANALISIS HASIL

5.1 Skenario Pengujian

Pengujian yang dilakukan meliputi dua,tahap yaitu tahap pertama dengan menggunakan metode KNN untuk mengklasifikasi data yang didalamnya terdapat data dengan *missing value* dengan mencari nilai K terbaik .Dan tahap kedua pengujian ID3 dengan data yang mengandung *missing value* dan data yang tidak mengandung *missing value*

5.1.1 Skenario Pengujian Menggunakan K-NN.

Pengujian dengan menggunakan metode K-NN ini adalah salah satu metode yang digunakan untuk masalah missing data. Metode K-NN adalah menghitung jarak terdekat dari data yang missing value ke training sample untuk menentukan KNN. Dalam 100 data rekam medik ada 7 data yang mengandung nilai *missing value* di dalam masing-masing atribut yang berbeda.Berikut data yang mengandung *missing value* pada Tabel 5.1.

Tabel 5.1 Data missing value

Pasien	Gejala								Typus
	D	P	L	IT	PTE	LTK	UJM	N	
P58	-	x	X	X	V	x	V	v	Positif
P60	V	V	X	V	V	X	-	V	positif
P67	V	X	X	X	X	X	X	-	Negatif
P69	-	X	X	V	X	X	V	V	Positif
P79	V	-	X	X	V	X	X	X	Positif
P83	V	V	X	-	V	X	X	V	negatif
P87	x	V	v	X	v	X	-	X	positif



Skenario pengujian pertama dalam K-NN ini menggunakan cara menghilangkan salah satu dari atribut data yang tidak mengandung nilai *miss value* atau data yang lengkap ,sebagai uji coba untuk mencari nilai K terbaik yang akan digunakan untuk pengujian selanjutnya.uji coba ini menggunakan 5 data dan salah satu atribut yaitu demam dihilangkan,Kemudian diuji dengan 90 data latih,untuk mendapatkan nilai K terbaik . Berikut pada Tabel 5.2 pengujian KNN dengan menghilangkan atribut demam.

Tabel 5.2 Pengujian KNN

Pasien	Gejala								Typus
	D	P	L	IT	PTE	LTK	UJM	N	
P19		v	x	V	x	x	x	v	Positif
P22		x	v	X	x	x	v	x	Negatif
P46		v	v	V	v	x	v	v	Positif
P75		v	x	V	v	v	v	x	Negatif
P99		v	x	V	x	x	x	v	Positif

5.1.2 Skenario pengujian KNN pertama untuk mencari nilai K

Dari hasil pengujian yang telah dilakukan, didapatkan hasil kemiripan dan jarak K terdekat adalah 1 dengan akurasi terbesar yaitu 100% . Berikut Pada Tabel 5.3 akan ditunjukkan hasil 4 pengujian nilai K dengan menggunakan 5 data uji dengan 90 data latih.

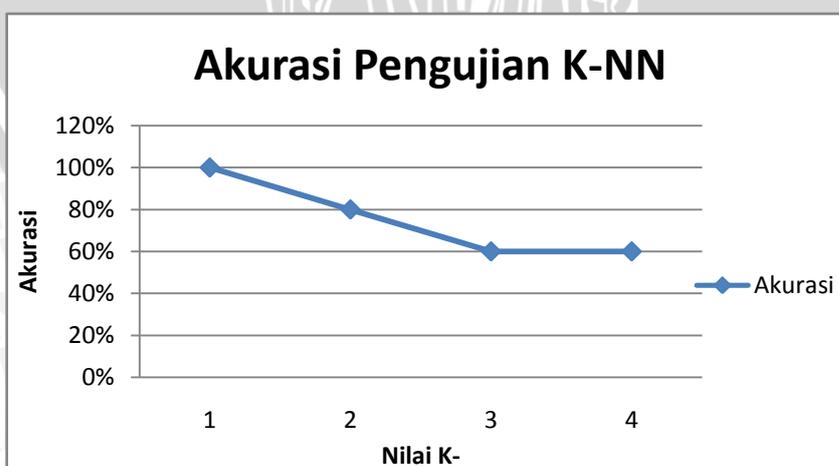
Tabel 5.3 Skenario Pengujian KNN

Nilai K	Data ke-	Prediksi	Nilai sebenarnya	Akurasi
1	P19	V	V	100%
	P22	V	V	
	P46	X	X	
	P75	X	X	

	P99	V	v	
2	P19	V	V	80%
	P22	V	V	
	P46	V	X	
	P75	X	X	
	P99	V	v	
3	P19	V	V	60%
	P22	V	V	
	P46	V	X	
	P75	V	X	
	P99	V	v	
4	P19	V	V	60%
	P22	V	V	
	P46	V	X	
	P75	V	X	
	P99	V	v	

5.1.3 Hasil Pengujian K-NN Pertama

Dari hasil pengujian yang telah dilakukan, didapatkan hasil akurasi yang beragam. Pada Gambar 5.1 akan ditunjukkan hasil perbandingan akurasi rata - rata dari 4 pengujian nilai K dengan menggunakan 5 data uji dengan 90 data latih.



Gambar 5.1 Grafik Perbandingan Akurasi Pengujian K-NN.

Seperti yang ditunjukkan pada Gambar 5.1 maka dapat dilihat dari 4 pengujian nilai K dengan menggunakan 5 data uji dengan 90 data latih. K1 mempunyai akurasi tertinggi yaitu 100%, dan akurasi terendah K3 dan K4 yaitu 60%.

5.1.4 Skenario Pengujian KNN kedua Menggunakan 7 data missing value.

Dalam pengujian K-NN uji coba pertama telah ditemukan nilai K terbaik dengan akurasi 100% adalah K=1, yang akan digunakan untuk pengujian 7 data *missing value* selanjutnya. Berikut Tabel 5.4 merupakan scenario pengujian K-NN.

Tabel 5.4 Skenario Pengujian 7 data missing value

Data Missing value	Pengujian	Hasil KNN
P58	80 Data latih	P23
	70 Data latih	P70
	60 Data latih	P23
P60	80 Data latih	P5
	70 Data latih	P83
	60 Data latih	P5
P67	80 Data latih	P44
	70 Data latih	P76
	60 Data latih	P44
P69	80 Data latih	P4
	70 Data latih	P23
	60 Data latih	P4
P79	80 Data latih	P38
	70 Data latih	P39
	60 Data latih	P38
P83	80 Data latih	P60



P87	70 Data latih	P60
	60 Data latih	P60
	80 Data latih	P53
P87	70 Data latih	P100
	60 Data latih	P55

Nilai *Missing value* pada data P58 setelah diuji dengan 80 data latih memiliki nilai kemiripan dan jarak terdekat dengan data P23, uji coba ke dua dengan 70 data latih, nilai kemiripan dan jarak terdekat dengan data yang sama yaitu P70, dan uji coba ke tiga diuji dengan 60 data latih nilai kemiripan dan jarak terdekat yaitu P23.

5.2 Skenario Pengujian ID3

Pengujian dilakukan untuk menentukan tingkat akurasi dari sistem yang telah dibuat, sehingga akan diketahui program dapat digunakan untuk mendiagnosa awal penyakit thypus dengan baik. Akurasi dihitung menggunakan persamaan (5-1):

$$Akurasi = \frac{Jumlah\ data\ benar}{Jumlah\ data} \times 100\% \dots\dots\dots(5-1)$$

Pengujian yang dilakukan meliputi dua tahap pengujian ID3, tahap pertama yaitu dengan menggunakan data yang mengandung *missing value* dan data yang tidak mengandung *missing value*. Pada skenario pengujian ID3 dilakukan dengan 3 pengujian dengan 10 data uji yang sama dan data latih yang berbeda. Rangkaian skenario pengujian ini dilakukan untuk mengetahui tingkat akurasi dari sistem.

5.2.1 Skenario Pengujian ID3 Menggunakan 90 data

Pada tahap pengujian ini system akan diuji dengan menggunakan 90 data yang didalamnya terdapat beberapa data dengan *missing value*. Pengujian menggunakan 10 data uji yang tetap dengan data latih yang diambil secara acak pada masing – masing perbandingan. Pada Tabel 5.5 berikut akan ditunjukkan skenario pengujian yang telah dibuat.

Tabel 5.5 Skenario Pengujian

Perbandingan	Pengujian	Akurasi	Akurasi Rata- Rata
90:10	1.	80%	80%
	2.	80%	
80:10	1.	80%	84%
	2.	80%	
	3.	90%	
	4.	90%	
	5.	80%	
70:10	1.	90%	82%
	2.	80%	
	3.	90%	
	4.	70%	
	5.	80%	
60:10	1.	70%	80%
	2.	80%	
	3.	80%	
	4.	90%	
	5.	80%	

Dalam pengujian tingkat akurasi akan dilakukan perbandingan antara kelas hasil diagnosa yang dihasilkan sistem dengan kelas diagnosa pada data asli. Perhitungan dengan ID3 dilakukan untuk membentuk pohon aturan yang selanjutnya disimpan sebagai pengetahuan sistem dalam menyelesaikan kasus yang akan datang.

5.2.1.1 Pengujian 90:10

Pada skenario pengujian 90 : 10, menggunakan 10 data uji yang tetap untuk masing – masing pengujian dan 90 data latih.

Pengujian pada skenario ini hanya dilakukan satu kali karena jumlah data yang terbatas .Persentase akurasi menggunakan 90 data latih dan 10 data uji dari total 90 data adalah 80%..

5.2.1.2 Pengujian 80:10

Pada skenario pengujian 80 : 10, menggunakan 10 data uji yang tetap untuk masing – masing pengujian dan 80 data latih yang digunakan diambil secara acak dan dilakukan pengujian sebanyak 5 kali. Persentase akurasi menggunakan 80 data latih dan 10 data uji dari total 90 data. Didapat nilai akurasi paling rendah adalah 80.00% sedangkan nilai akurasi terbaik adalah 90.00%.

5.2.1.3 Pengujian 70:10

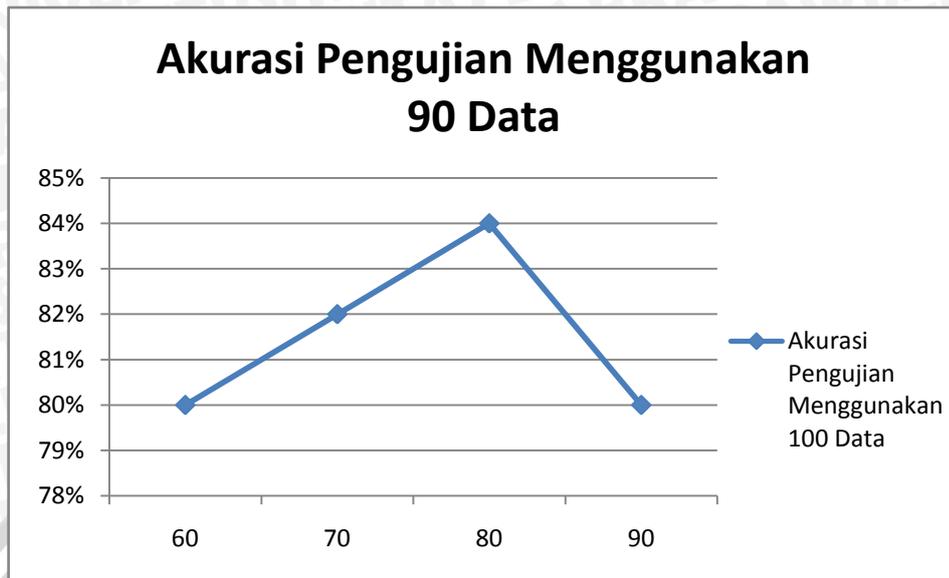
Pada skenario pengujian 70 : 10, menggunakan 10 data uji yang tetap untuk masing – masing pengujian, dan 70 data latih yang digunakan diambil secara acak dan dilakukan pengujian sebanyak 5 kali. Persentase akurasi menggunakan 70 data latih dan 10 data uji dari total 90 data. Didapat nilai akurasi paling rendah adalah 70.00% sedangkan nilai akurasi terbaik adalah 90.00%.

5.2.1.4 Pengujian 60:10

Pada skenario pengujian 60 : 10, menggunakan 10 data uji yang tetap untuk masing – masing pengujian, dan 60 data latih yang digunakan diambil secara acak dan dilakukan pengujian sebanyak 5 kali. Persentase akurasi menggunakan 60 data latih dan 10 data uji dari total 90 data. Didapat nilai akurasi paling rendah adalah 70.00% sedangkan nilai akurasi terbaik adalah 90.00%.

5.2.1.5 Hasil Pengujian dengan 90 Data

Dari hasil pengujian yang telah dilakukan, didapatkan hasil akurasi yang beragam. Pada Gambar 5.2 akan ditunjukkan hasil perbandingan akurasi rata - rata dari 3 pengujian yang dilakukan pertama menggunakan 90 data.



Gambar 5.2 Grafik Perbandingan Akurasi Rata –Rata

Seperti yang ditunjukkan pada Gambar 5.2 maka dapat dilihat dari pengujian 90 data bahwa akurasi rata – rata terendah yaitu 80% pada pengujian menggunakan 60 data latih dan akurasi rata – rata tertinggi 84% menggunakan 80 data latih.

5.2.2 Skenario Pengujian ID3 menggunakan 80 data

Pada tahap pengujian ini system akan diuji dengan menggunakan 80 data yang didalamnya tidak terdapat nilai *missing value* didalamnya. Pengujian menggunakan 10 data uji yang tetap dengan data latih yang diambil secara acak pada masing – masing perbandingan. Pada Tabel 5.6 berikut akan ditunjukkan skenario pengujian yang telah dibuat.

Tabel 5.6 Skenario Pengujian

Perbandingan	Pengujian	Akurasi	Akurasi Rata- Rata
80:10	1.	80%	86%
	2.	90%	
	3.	80%	
	4.	90%	
	5.	90%	

70:10	1.	80%	84%
	2.	80%	
	3.	80%	
	4.	90%	
	5.	90%	
60:10	1.	80%	82%
	2.	80%	
	3.	90%	
	4.	80%	
	5.	80%	

Dalam pengujian tingkat akurasi akan dilakukan perbandingan antara kelas hasil diagnosa yang dihasilkan sistem dengan kelas diagnosa pada data asli. Perhitungan dengan ID3 dilakukan untuk membentuk pohon aturan yang selanjutnya disimpan sebagai pengetahuan sistem dalam menyelesaikan kasus yang akan datang.

5.2.2.1 Pengujian 80:10

Pada skenario pengujian 80 : 10, menggunakan 10 data uji yang tetap untuk masing – masing pengujian, dan 80 data latih yang digunakan diambil secara acak dan dilakukan pengujian sebanyak 5 kali.

Persentase akurasi menggunakan 80 data latih dan 10 data uji dari total 80 data. Didapat nilai akurasi paling rendah adalah 80.00% sedangkan nilai akurasi terbaik adalah 90.00%.

5.2.2.2 Pengujian 70:10

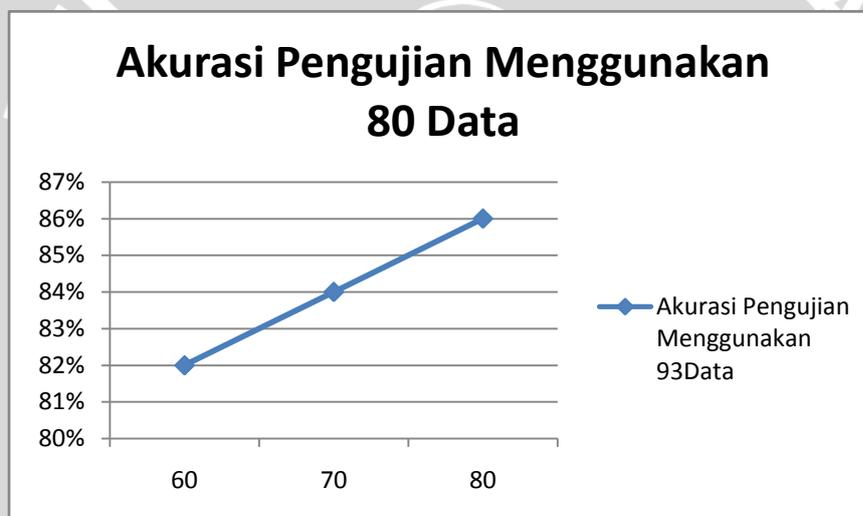
Pada skenario pengujian 70 : 10, menggunakan 10 data uji yang tetap untuk masing – masing pengujian, dan 70 data latih yang digunakan diambil secara acak dan dilakukan pengujian sebanyak 5 kali. Persentase akurasi menggunakan 70 data latih dan 10 data uji dari total 80 data. Didapat nilai akurasi paling rendah adalah 80.00% sedangkan nilai akurasi terbaik adalah 90.00%.

5.2.2.3 Pengujian 60:10

Pada skenario pengujian 60 : 10, menggunakan 10 data uji yang tetap untuk masing – masing pengujian, dan 60 data latih yang digunakan diambil secara acak dan dilakukan pengujian sebanyak 5 kali. Persentase akurasi menggunakan 60 data latih dan 10 data uji dari total 80 data. Didapat nilai akurasi paling rendah adalah 80.00% sedangkan nilai akurasi terbaik adalah 90.00%.

5.2.2.4 Hasil Pengujian Berdasarkan 80 Data

Dari hasil pengujian yang telah dilakukan, didapatkan hasil akurasi yang beragam. Pada gambar 5.3 akan ditunjukkan hasil perbandingan akurasi rata - rata dari 3 pengujian yang dilakukan menggunakan 80 data.

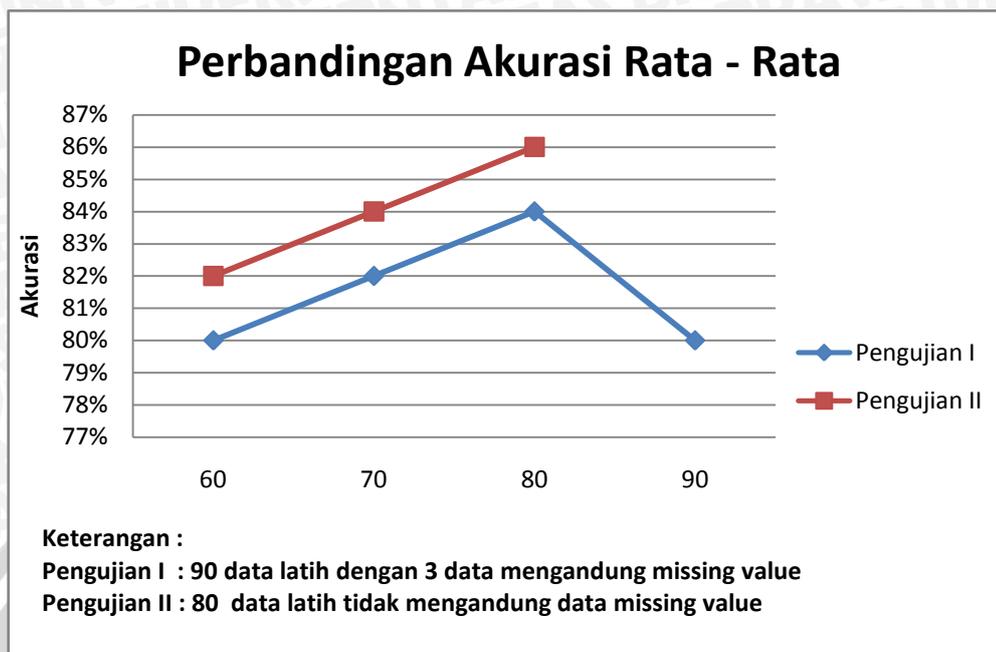


Gambar 5.3 Grafik Perbandingan Akurasi Rata –Rata

Seperti yang ditunjukkan pada gambar 5.3 pada pengujian menggunakan 80 data akurasi rata – rata terendah 82% dengan 60 data latih dan akurasi tertinggi yaitu 86% pada pengujian menggunakan 80 data latih.

5.3. Hasil Pengujian Berdasarkan Jumlah Data Latih

Dari hasil pengujian yang telah dilakukan, didapatkan hasil akurasi yang beragam. Pada Gambar 5.4 akan ditunjukkan hasil perbandingan akurasi rata - rata dari 3 pengujian yang dilakukan pertama menggunakan 90 data dan pengujian kedua menggunakan 80 data.



Gambar 5.4 Grafik Perbandingan Akurasi Rata –Rata

Seperti yang ditunjukkan pada Gambar 5.4 maka dapat dilihat dari pengujian 90 data bahwa akurasi rata – rata terendah yaitu 80% pada pengujian menggunakan 60 data latih dan akurasi rata – rata tertinggi 84% menggunakan 80 data latih. Sedangkan pada pengujian menggunakan 80 data akurasi rata – rata terendah 82% dengan 60 data latih dan akurasi tertinggi yaitu 86% pada pengujian menggunakan 80 data latih.

5.4 Analisa Hasil Pengujian

Data yang digunakan berasal dari Rumah Sakit Umum Besuki berjumlah 100 data rekam medis dengan dua diagnosa yaitu positif dan dan negatif penyakit thypus, pada 100 data rekam medis ini terdapat beberapa data yang *missing value*.. dengan disertai 8 kriteria gejala . Pada dataset terdapat kelas yang berbeda pada data yang demam, pusing, lesu, infeksi tenggorokan, perut tidak enak, lidah tengah kotor, ujung lidah merah, dan ngigau. Sehingga menyebabkan sistem membentuk pohon aturan yang menghasilkan dua kelas untuk solusi. Hal tersebut dapat diatasi dengan bantuan informasi dari pakar atau dokter, yaitu ketika sistem memberikan prediksi diagnosa dengan dua kelas, maka solusi yang dapat

diberikan yaitu dengan cek laboratorium. Contoh pada kasus yang diprediksi sistem memiliki diagnosa positif dan negatif maka diagnosa yang diberikan untuk kasus tersebut adalah cek laboratorium. Untuk mempermudah maka ditambahkan kelas baru cek lab, dengan mengubah data yang sama tetapi memiliki diagnosa yang berbeda maka diagnosa dirubah menjadi cek lab.

Missing value ditangani menggunakan K-NN, metode K-NN ini adalah mencari nilai kemiripan dan jarak K. Setelah diuji coba didapat nilai K adalah 1, dengan akurasi 100% pada gambar 5.1 yang digunakan untuk uji coba selanjutnya.

Sistem dapat menjalankan metode ID3 pada data yang mengandung *missing value* dan data yang tidak mengandung *missing value* berjalan dengan cukup baik dan sesuai dengan perhitungan manualnya.

Pada uji coba pertama menggunakan 90 data yang didalamnya terdapat nilai *missing value* menunjukkan nilai rata-rata akurasi tertinggi yaitu 84% dengan 80 data latih, dan akurasi terendah 80% menggunakan 60 data latih. Uji coba kedua menggunakan 80 data yang didalamnya tidak ada *missing value*, menunjukkan nilai rata-rata tertinggi didapat pada pengujian menggunakan 80 data latih yaitu 86.00%, dan terendah 82.00% menggunakan 60 data latih seperti yang ditunjukkan pada gambar 5.4.

Akurasi tertinggi berbeda dengan akurasi rata-rata dikarenakan pengambilan data yang acak sehingga pada setiap skenario akan mendapat hasil data yang tidak diduga dan data akurasi juga menjadi sangat beragam.

Kualitas dataset yang mengandung *missing value* dan dataset yang tidak mengandung *missing value*, juga mempengaruhi tingkat akurasi yang dihasilkan karena dataset *missing value* tidak bisa memutuskan hasil diagnosa benar secara pasti, contoh data yang mengandung *missing value* tidak semua hasil akhirnya merupakan cek lab, hasil cek lab apabila data dengan kriteria sama dengan hasil diagnosa berbeda, kemudian barulah system akan memutuskan data *missing value* tersebut menjadi cek lab. Dengan hasil yang mengandung untuk membentuk pengetahuan sistem dalam menyelesaikan kasus baru.