

## BAB V

### PENGUJIAN DAN ANALISIS

#### 5.1 Proses Pengujian

Berdasarkan penjelasan pada bab sebelumnya, bahwa proses pengujian terhadap sistem klasifikasi kondisi penderita penyakit hepatitis dengan menggunakan metode *Support Vector Machine* yaitu dengan menggunakan dataset *Hepatitis Domain*. Dataset ini kemudian akan digunakan sebagian sebagai data latih dan sebagian lagi sebagai data uji. Data latih yang digunakan diambil secara random dari dataset sebanyak 20,40,60, 80 *record* data, sedangkan untuk data latih diambil secara random dari dataset sebanyak 75 *record* data. Kemudian untuk parameter yang digunakan pada proses pengujian yaitu menggunakan nilai *C* (*Complexity*) yang merupakan parameter user pada proses *training* SVM. Proses pengujian dengan menggunakan SVM, akan digunakan nilai *C* yaitu 10, 20, 30, 40, 50,.... 100. Skenario pengujian yang dilakukan adalah dengan melakukan pelatihan terhadap data latih dengan menggunakan nilai *C* yang sudah ditentukan untuk menguji data uji sebanyak 75 *record* data, jadi pengujian akan dilakukan sebanyak 10 kali berdasarkan jumlah nilai *C* yang ada. Tujuan dari skenario ini adalah untuk mencari nilai akurasi terbaik pada proses pengujian, dan mendapatkan nilai *C* yang memberikan akurasi terbaik terhadap proses klasifikasi oleh sistem. Pada proses ini juga akan diukur nilai *sensitivity* dan *spesitivity* pada sistem, untuk mengetahui tingkat pengenalan sistem terhadap kelas negatif dan kelas positif pada sistem. Tujuannya adalah untuk mengetahui kecenderungan sistem dalam mengklasifikasikan dataset yang digunakan.

#### 5.2 Hasil Pengujian

Berdasarkan skenario pengujian, berikut ini akan ditampilkan hasil pengujian untuk mendapatkan nilai akurasi, nilai *sensitivity*, dan *spesitivity*. Proses pengujian dilakukan sesuai dengan skenario, namun untuk jumlah atribut data yang digunakan akan berbeda, dataset yang digunakan yaitu dataset dengan 19 atribut dan dataset dengan 15 atribut. Hal ini dilakukan karena pada dataset

Hepatitis Domain, terdapat data dengan *missing value*, maka kemudian dilakukan pengujian untuk mengetahui tingkat akurasi dengan 19 atribut dan 15 atribut pada dataset.

### 5.2.1 Hasil Confusion Table

Untuk proses pengukuran nilai akurasi, *sensitivity*, dan *spesitiviti*, maka digunakan *confusion table* untuk mempermudah pengukuran, penjelasan mengenai *confusion table* telah dibahas pada Bab 2. Berikut ini adalah hasil dari *confusion table* berdasarkan jumlah data latih dan jumlah atribut data yang digunakan.

1. *Confusion table* dengan 19 atribut data

**Tabel 5.1** *Confusion table* Datalatih 20

Datalatih = 20				
C	Confusion Table			
	TP	FP	TN	FN
10	61	13	0	1
20	62	13	0	0
30	62	13	0	0
40	61	13	0	1
50	61	13	0	1
60	61	13	0	1
70	61	13	0	1
80	61	13	0	1
90	61	13	0	1
100	61	13	0	1

**Tabel 5.2** *Confusion table* Datalatih 40

Datalatih = 40				
C	Confusion Table			
	TP	FP	TN	FN
10	62	13	0	0
20	62	13	0	0
30	62	12	1	0
40	62	13	0	0
50	62	13	0	0
60	62	13	0	0

70	62	13	0	0
80	62	13	0	0
90	62	13	0	0
100	62	13	0	0

**Tabel 5.3** *Confusion table* Datalatih 60

Datalatih = 60				
C	Confusion Table			
	TP	FP	TN	FN
10	62	13	0	0
20	62	13	0	0
30	62	13	0	0
40	62	13	0	0
50	62	13	0	0
60	62	13	0	0
70	62	13	0	0
80	62	13	0	0
90	62	13	0	0
100	62	13	0	0

**Tabel 5.4** *Confusion table* Datalatih 80

Datalatih = 80				
C	Confusion Table			
	TP	FP	TN	FN
10	62	13	0	0
20	62	13	0	0
30	62	12	1	0
40	62	13	0	0
50	62	13	0	0
60	62	13	0	0
70	62	13	0	0
80	62	13	0	0
90	62	13	0	0
100	62	13	0	0



2. *Confusion table* dengan 15 atribut data**Tabel 5.5** *Confusion table* Datalatih 20

Datalatih = 20				
C	Confusion Table			
	TP	FP	TN	FN
10	62	13	0	0
20	62	13	0	0
30	61	13	0	1
40	61	13	0	1
50	61	13	0	1
60	61	13	0	1
70	61	13	0	1
80	61	13	0	1
90	61	13	0	1
100	61	13	0	1

**Tabel 5.6** *Confusion table* Datalatih 40

Datalatih = 40				
C	Confusion Table			
	TP	FP	TN	FN
10	61	12	2	1
20	60	7	6	2
30	60	7	6	2
40	60	10	3	2
50	61	11	2	1
60	62	12	1	0
70	62	12	1	0
80	62	13	0	0
90	62	13	0	0
100	62	13	0	0

**Tabel 5.7** *Confusion table* Datalatih 60

Datalatih = 60				
C	Confusion Table			
	TP	FP	TN	FN
10	60	9	4	2
20	60	8	5	2
30	60	9	4	2
40	60	10	3	2
50	60	10	3	2

60	61	11	2	1
70	60	10	3	2
80	60	10	3	2
90	60	9	4	2
100	60	8	5	2

**Tabel 5.8** Confusion table Datalatih 80

Datalatih = 80				
C	Confusion Table			
	TP	FP	TN	FN
10	62	13	0	0
20	62	13	0	0
30	61	12	1	1
40	62	12	1	0
50	62	12	1	0
60	62	12	1	0
70	62	13	0	0
80	62	13	0	0
90	62	13	0	0
100	62	13	0	0

**5.2.2 Hasil Pengukuran Akurasi**

Setelah didapatkan nilai *confusion table* maka kemudian akan dilakukan proses nilai akurasi dari proses klasifikasi dengan menggunakan metode SVM. Untuk proses pengukuran tingkat akurasi dari sistem akan mengimplementasikan persamaan 2.24. Berikut ini adalah hasil pengukuran tingkat akurasi berdasarkan data latih yang digunakan dan jumlah atributnya.

1. Akurasi dengan 19 atribut data

**Tabel 5.9** Tingkat Akurasi dengan 19 atribut data (%)

C	Dataset			
	20	40	60	80
10	81.33	82.667	82.667	82.667
20	82.667	82.667	82.667	82.667
30	82.667	84	82.667	84
40	81.33	82.667	82.667	82.667
50	81.33	82.667	82.667	82.667
60	81.33	82.667	82.667	82.667
70	81.33	82.667	82.667	82.667



80	81.33	82.667	82.667	82.667
90	81.33	82.667	82.667	82.667
100	81.33	82.667	82.667	82.667
<b>Rata -Rata</b>	<b>81.5974</b>	<b>82.8003</b>	<b>82.667</b>	<b>82.8003</b>

## 2. Akurasi dengan 15 atribut data

**Tabel 5.10** Tingkat Akurasi dengan 15 atribut data (%)

C	Dataset			
	20	40	60	80
10	82.667	84	85.33	82.667
20	82.667	88	86.67	82.667
30	81.33	88	85.33	82.667
40	81.33	84	84	84
50	81.33	84	84	84
60	81.33	84	84	84
70	81.33	84	84	82.667
80	81.33	82.667	84	82.667
90	81.33	82.667	85.33	82.667
100	81.33	82.667	86.67	82.667
<b>Rata -Rata</b>	<b>81.5974</b>	<b>84.4001</b>	<b>84.933</b>	<b>83.0669</b>

Berdasarkan tabel 5.9 ditunjukkan bahwa pada data latih = 20, nilai rata-rata akurasi yang didapatkan adalah 81.5974 %, dimana akurasi tertinggi terletak pada nilai C = 20 dan 30 , dengan akurasi mencapai 82.667 %. Kemudian untuk nilai akurasi terendah didapatkan 81.33 %. Pada data latih = 40 nilai akurasi tertinggi adalah 84 % terletak pada nilai C = 30, serta akurasi terendah yang didapatkan adalah 82.667 % .Selanjutnya pada data latih = 60, didapatkan nilai akurasi tertinggi 82.667 % dengan rata – rata akurasi adalah 82.667 %. Pada data latih = 80, didapatkan nilai akurasi tertinggi 84 %, yang terletak pada nilai C = 30, dan akurasi terendah adalah 82.667 % dengan rata – rata tiap akurasi adalah 82.803 % .

Berdasarkan tabel 5.10 dapat dilihat bahwa pada data latih = 20 nilai akurasi tertinggi yang didapatkan adalah 82.667 % pada nilai C = 10,20, dan akurasi terendah yang didapatkan adalah 81.34 %. Kemudian untuk data latih = 40, nilai akurasi yang tertinggi didapatkan adalah 88 % pada nilai C = 20,30, dan nilai akurasi terendah yaitu 82.67 %. Pada data latih = 60, nilai akurasi tertinggi

terletak pada nilai  $C = 20$  dengan akurasi 86.67 %, dan untuk nilai akurasi terendah adalah 84. %. Selanjutnya untuk data latih = 80, nilai akurasi tertinggi terletak yaitu 84 % pada nilai  $C = 40,50,60$ , dan nilai akurasi terendah adalah 82.667 % dimana rata – rata akurasi yang diperoleh adalah 83.069 %.

### 5.2.3 Hasil pengukuran *Sensitivity* dan *Spesitivity*

*Sensitivity* adalah proporsi kasus dengan kelas tingkat positif yang diklasifikasikan sebagai kelas positif, sedangkan *specificity* adalah proporsi dengan kelas tingkat negatif diklasifikasikan sebagai kelas negatif. Tujuan dari pengukuran ini adalah untuk mengetahui tingkat pengenalan kelas negatif dan positif oleh sistem. Proses untuk mendapatkan hasil pengukuran nilai *sensitivity* dan nilai *specificity* dengan mengimplementasikan persamaan 2.25 dan 2.26. Berikut ini adalah hasil pengukuran tingkat *sensitivity* dan *spesitivity* berdasarkan jumlah atribut data.

1. Tingkat *sensitivity* dan *spesitivity* 19 atribut data

**Tabel 5.11** *Sensitivity* 19 atribut (%)

C	<i>Sensitivity</i>			
	20	40	60	80
10	98.4	100	100	100
20	100	100	100	100
30	100	100	100	100
40	98.4	100	100	100
50	98.4	100	100	100
60	98.4	100	100	100
70	98.4	100	100	100
80	98.4	100	100	100
90	98.4	100	100	100
100	98.4	100	100	100

**Tabel 5.12** *Spesitivity* 19 atribut (%)

C	<i>Spesitivity</i>			
	20	40	60	80
10	0	0	0	0
20	0	0	0	0
30	0	7.7	0	7.7
40	0	0	0	0

50	0	0	0	0
60	0	0	0	0
70	0	0	0	0
80	0	0	0	0
90	0	0	0	0
100	0	0	0	0

2. Tingkat *sensitivity* dan *specificity* 15 atribut data

**Tabel 5.13** *Sensitivity* 15 atribut (%)

C	<i>Sensitivity</i>			
	20	40	60	80
10	100	98.4	96.8	100
20	100	96.8	96.8	100
30	98.4	96.8	96.8	98.4
40	98.4	96.8	96.8	100
50	98.4	98.4	96.8	100
60	98.4	100	98.4	100
70	98.4	100	96.8	100
80	98.4	100	96.8	100
90	98.4	100	96.8	100
100	98.4	100	96.8	100

**Tabel 5.14** *Spesitivity* 15 atribut (%)

C	<i>Spesitivity</i>			
	20	40	60	80
10	0	15.4	30.8	0
20	0	46.2	38.5	0
30	0	46.2	30.8	7.7
40	0	23.1	23.1	7.7
50	0	15.4	23.1	7.7
60	0	7.7	15.4	7.7
70	0	7.7	23.1	0
80	0	0	23.1	0
90	0	0	30.8	0
100	0	0	38.5	0

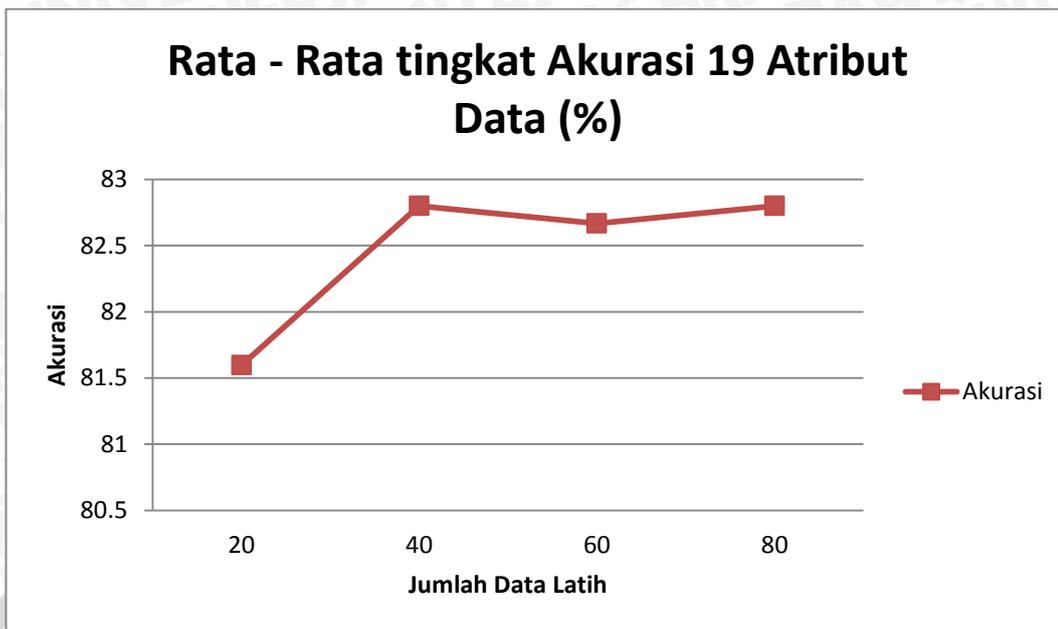
Tabel - tabel di atas merupakan hasil dari pengukuran *sensitivity* dan *spesitivity* dari data latih berdasarkan nilai C (*Complexity*). Pada data latih dengan 19 atribut diperoleh bahwa nilai *spesitivity* tertinggi adalah 7.7 % dan nilai



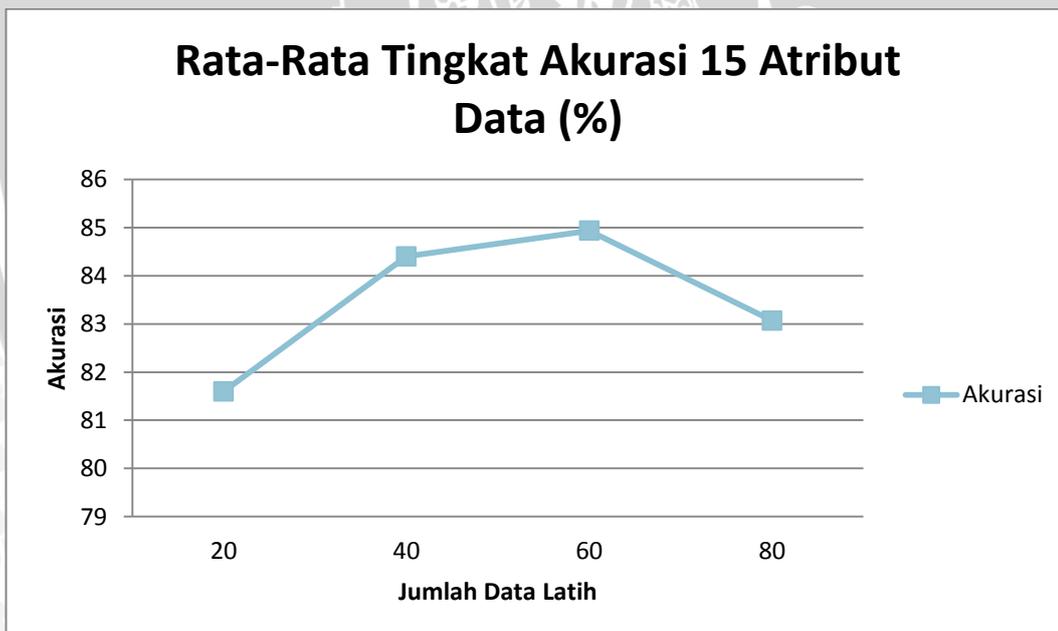
spesitivity terendah adalah 0 %. Sedangkan untuk nilai *sensitivity* diperoleh bahwa nilai tertinggi adalah 100 % dan nilai terendah adalah 98.4 %. Kemudian untuk data latih dengan 15 atribut diperoleh nilai *spesitivity* tertinggi adalah 46.2 % dan *spesitivity* terendah adalah 0 %. Selanjutnya untuk nilai *sensitivity* tertinggi adalah 100 % dan terendah adalah 96.8 %. Berdasarkan tabel pengukuran di atas dilihat bahwa, dalam proses pengukuran nilai *sensitivity* dan *spesitivity* menunjukkan bahwa nilai *sensitivity* lebih tinggi daripada nilai *spesitivity*, hal ini berarti sistem lebih mengenali kelas yang positif dibandingkan dengan kelas yang negatif. Berdasarkan analisis yang ada, hal tersebut terjadi karena pada data uji dan data latih, prosentase nilai kelas positif jauh lebih banyak dibandingkan prosentasi nilai kelas negatif, sehingga dalam proses pelatihan dengan metode SVM, sistem memiliki kecenderungan untuk mengenali dan mengklasifikasikan kelas yang positif.

### 5.3 Analisa Hasil

Berikut ini adalah akan dijelaskan mengenai analisa hasil yang didapatkan berdasarkan hasil pengukuran tingkat akurasi, tingkat *sensitivity* dan tingkat *spesitivity* pada proses klasifikasi dengan menggunakan metode *Support Vector Machine*. Proses pelatihan pada SVM adalah dengan mengimplementasikan persamaan primal SVM yaitu pada persamaan 2.9, dimana pada persamaan tersebut terdapat parameter  $C$  (*Complexity*) yang merupakan parameter user yang diinputkan untuk mencari nilai  $w$  (*weight*) dan nilai  $b$  (*bias*). Pada bab sebelumnya telah dijelaskan bahwa rentang nilai  $C$  adalah  $0 < C < \infty$ . Nilai  $C$  digunakan untuk memperkecil *error* pada proses training dengan SVM, dimana jika nilai  $C$  semakin besar, maka nilai *error* pada klasifikasi akan kecil dan sebaliknya. Berdasarkan hasil pengukuran tingkat akurasi yang didapatkan pada dataset dengan jumlah atribut 19 dan jumlah atribut 15 dapat dibuat grafik tingkat akurasi. Adapun grafik tingkat akurasi dapat dilihat pada gambar 5.1 dan 5.2.



Gambar 5.1 Grafik tingkat Akurasi (19 atribut)



Gambar 5.2 Grafik tingkat Akurasi (15 atribut)

Dari data pengujian yang dilakukan di atas, maka akan didapatkan nilai perubahan akurasi terhadap nilai parameter yang diberikan, sehingga setiap nilai parameter dapat mengakibatkan perubahan nilai akurasi pada data uji. Pada hasil pengujian terdapat nilai – nilai yang berpengaruh terhadap akurasi. Peningkatan parameter C mengakibatkan akurasi meningkat sekitar 2% sampai 8%, dimana

pada nilai C tertentu memberikan akurasi yang tertinggi pada proses klasifikasi, dan juga pada nilai C tertentu memberikan nilai akurasi terendah. Pada gambar 5.1 dapat dilihat bahwa akurasi tertinggi yang diperoleh adalah pada 60 data latih, sedangkan akurasi terendah pada 20 data latih. Jika dilihat bahwa tingkat akurasi mengalami peningkatan sesuai dengan jumlah data latih dan nilai C yang digunakan. Dari hasil yang diperoleh pada tabel akurasi sesuai dengan pengujian, rata – rata akurasi mengalami peningkatan sesuai dengan jumlah data latih yang dilakukan terhadap data uji.

Dari hasil pengujian yang dilakukan diperoleh bahwa rata – rata akurasi tertinggi yaitu 82.08 % pada data latih dengan jumlah atribut 19 dan 84.93 % pada data latih dengan jumlah atribut 15. Kemudian untuk parameter C yang memberikan akurasi tertinggi pada data latih dengan jumlah atribut 19 adalah nilai C = 30 dengan akurasi 84 %, dengan jumlah data latih 80. Sedangkan untuk data latih dengan jumlah atribut 15 didapatkan nilai akurasi tertinggi yaitu 88 pada nilai C = 40 dan jumlah data latih adalah 40.

Selanjutnya pada pengukuran *sensitivity* dan *spesitivity* pada data latih dengan jumlah atribut 19 dan jumlah atribut 15. Pada data latih dengan jumlah atribut 19 diperoleh nilai *sensitivity* maksimum adalah 100 % sedangkan nilai minimumnya adalah 98.4 %. Kemudian nilai *spesitivity* maksimum didapatkan adalah 7.7 % dan nilai minimum yang didapatkan adalah 0%. Pada data latih dengan jumlah atribut 15, diperoleh nilai *sensitivity* maksimum adalah 100%, sedangkan nilai minimumnya adalah 98.4 %. Selanjutnya untuk nilai *spesitivity* maksimum diperoleh 46.2 % , dan nilai minimumnya adalah 0 %. Dari hasil pengujian *sensitivity* dan *spesitivity*, dapat dilihat bahwa nilai akurasi *sensitivity* lebih besar daripada nilai *spesitivity*, hal ini menunjukkan bahwa pada data uji yang memiliki proporsi nilai kelas positif yang lebih banyak daripada kelas yang negatif. Dari sini dapat dilihat bahwa sistem lebih mengenali kelas yang positif dibandingkan kelas yang negatif, hal ini terjadi karena proporsi dataset memang memiliki jumlah kelas positif yang lebih banyak dari kelas negatif.

