

BAB IV

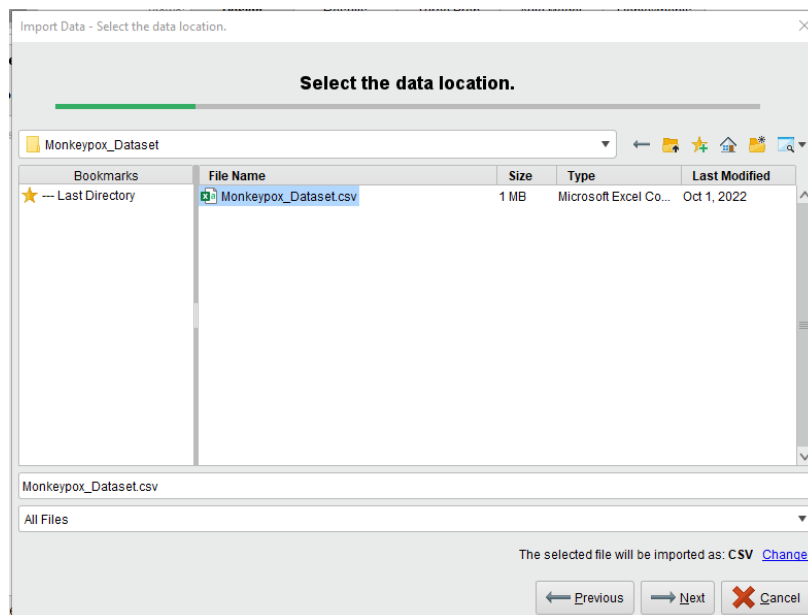
HASIL DAN PEMBAHASAN

4.1. Hasil dan Pembahasan

Pada penelitian ini digunakan seleksi fitur PSO dan Algoritma Decision Tree C4.5. topik yang diambil dalam penelitian ini adalah wabah penyakit monkeypox dengan jumlah data sebesar 25.000 data kemudian dilakukan pengolahan preprocessing sehingga data siap untuk dilakukan pemrosesan mining. Data sample ini yang akan dihitung probabilitasnya, kemudian akan dicari tingkat akurasi tertinggi dari data tersebut dan akan diimplementasikan pada tools RapidMiner.

4.1.1. Hasil Pengolahan Dataset

Pada hasil pengolahan dalam dataset adalah hasil dari pemrosesan saat dataset siap digunakan untuk dilakukan preprocessing. Tool yang digunakan adalah Rapidminer kemudian dataset yang digunakan merupakan data excel dengan format csv file. Pada tahapan ini dataset di import kedalam tool rapidminer seperti gambar berikut.



Gambar 4. 1 Proses Import Dataset

Setelah dipilih menu Next maka akan muncul permintaan untuk melihat spesifikasi pada format data yang akan dipilih. Pada tahapan ini digunakan untuk memastikan dataset range yang kita miliki pada setiap atribut agar sesuai untuk dilakukan preprocessing. Berikut adalah spesifikasi format data.

The screenshot shows a dialog box titled "Specify your data format" with the following settings:

- Header Row: 1
- Start Row: 1
- Column Separator: Comma ","
- File Encoding: windows-12...
- Escape Character: \
- Decimal Character: .
- Use Quotes: "
- Trim Lines
- Skip Comments: #

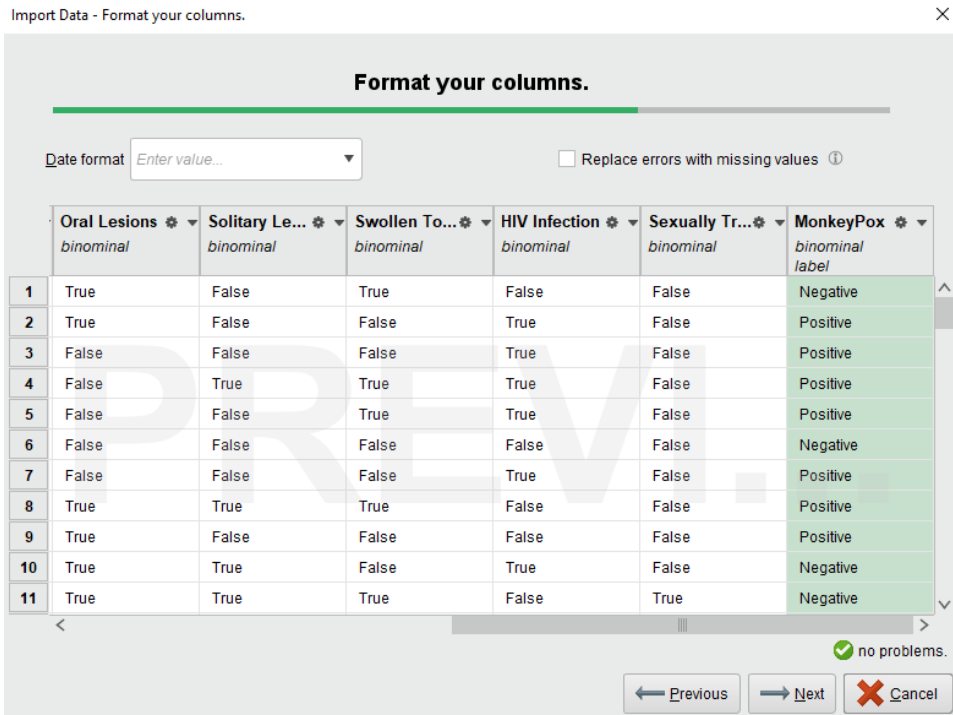
Below the settings is a data preview table:

	Patient_ID	Systemi...	Rectal P...	Sore Thr...	Penile O...	Oral Les...	Solitary ...	Swollen ...	HIV Infec...	Sexual
2	P0	None	False	True	True	True	False	True	False	False
3	P1	Fever	True	False	True	True	False	False	True	False
4	P2	Fever	False	True	True	False	False	False	True	False
5	P3	None	True	False	False	False	True	True	True	False
6	P4	Swollen ...	True	True	True	False	False	True	True	False
7	P5	Swollen ...	False	True	False	False	False	False	False	False
8	P6	Fever	False	True	False	False	False	False	True	False
9	P7	Fever	True	True	False	True	True	True	False	False
10	P8	Muscle A...	False	True	True	True	False	False	False	False

At the bottom right, there is a status "no problems." and buttons for "Previous", "Next", and "Cancel".

Gambar 4. 2 Format Spesifik Data

Setelah format data dilakukan maka akan muncul tampilan untuk dilakukan format pada kolom data pada dataset. Pada proses ini akan dilakukan perubahan tipe atribut sesuai dengan karakteristik atribut yang dimiliki, dimana 2 varian data dalam satu attribute akan digunakan tipe data binominal, lalu jika terdapat lebih dari 2 varian data dalam atribut maka akan digunakan tipe data polynominal pada atribut data. Kemudian pada tahapan ini akan ditentukan juga atribut yang digunakan sebagai labelnya dimana atribut Monkeypox akan digunakan sebagai label dalam dataset ini. Berikut adalah hasil dari format data pada kolom dan juga table keterangan dari tipe data dalam atribut di dataset monkeypox tersebut..



Gambar 4. 3 Format Kolom dan Pelabelan Data

Tabel 4. 1 Tipe Data Atribut

No	Nama Atribut	Tipe Data
1	Patient_ID	Polynomial
2	Systemic Illness	Polynomial
3	Rectal Pain	Binominal
4	Sore Throat	Binominal
5	Penile Oedema	Binominal
6	Oral Lesions	Binominal
7	Solitary Lesion	Binominal
8	Swollen Tonsils	Binominal
9	HIV Infection	Binominal
10	Sexually Transmitted Infection	Binominal
11	MonkeyPox	Binominal

Pada tampilan dataset yang sudah di import dapat dilihat dengan menarik langsung

hasil dari file yang di import ke dalam lembar kerja di dalam tool rapidminer seperti pada hasil gambar berikut.



Gambar 4. 4 Tampilan Dataset pada Rapidminer

Setelah dirunning dalam tool rapidminer maka akan muncul statistik dalam dataset yang telah di import sebelumnya. Dalam statistik dataset yang dijalankan akan terlihat berpakah missing data yang terdapat dalam dataset kemudian nilai negative dan positif pada monkeypox yang ada dalam dataset. Berikut adalah tampilan hasil gambarnya.

Name	Type	Missing	Statistics	Filter (11 / 11 attributes):
Label MonkeyPox	Binominal	0	Least Negative (9091)	Most Positive (15909)
Patient_ID	Polynominal	0	Least P9999 (1)	Most P0 (1)
Systemic Illness	Polynominal	0	Least Muscle A [...] in (6150)	Most Fever (6382)
Rectal Pain	Binominal	0	Least True (12345)	Most False (12655)
Sore Throat	Binominal	0	Least False (12446)	Most True (12554)
Penile Oedema	Binominal	0	Least False (12388)	Most True (12612)
Oral Lesions	Binominal	0	Least True (12486)	Most False (12514)

Gambar 4. 5 Tampilan Statistik Dataset

4.1.2. Visualisasi Model Data Attribute

a. Systemic Illness

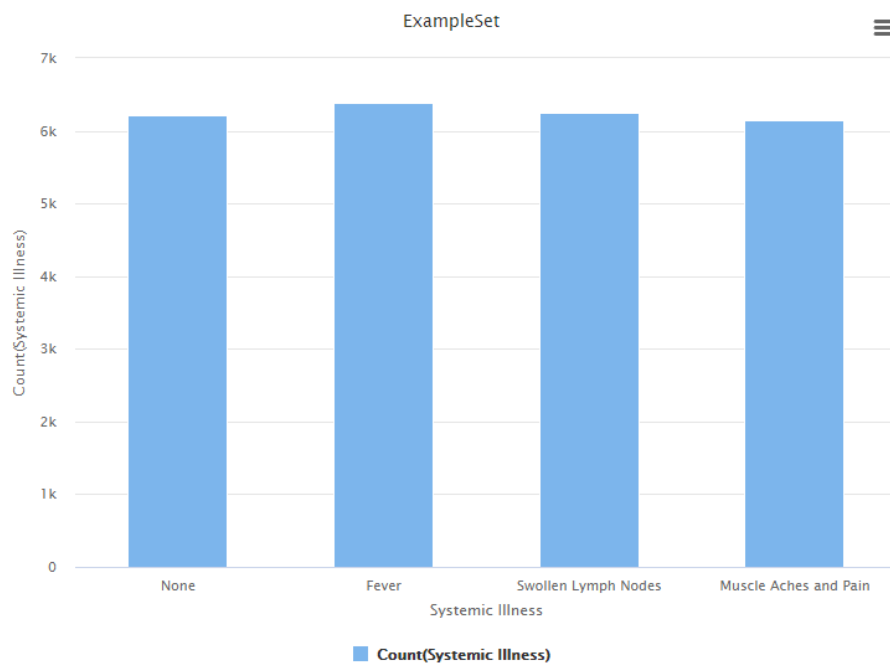
Systemic Illness merupakan penyakit sistemik yang merupakan gangguan yang dapat mempengaruhi beberapa organdan jaringan atau bahkan seluruh tubuh manusia. Data yang divisualisasikan pada gambar dibawah ini adalah

atribut dari Systemic Illness dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 2 Nominal Values pada Systemic Illness

No	Nominal Value	Absolute Count	Fraction
1	Fever	6382	0.255
2	Swollern Lymph Nodes	6252	0.250
3	None	6216	0.249
4	Muscle Aches and Pain	6150	0.246

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Systemic Illnes. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 6 Visualisasi Attribut Systemic Illnes

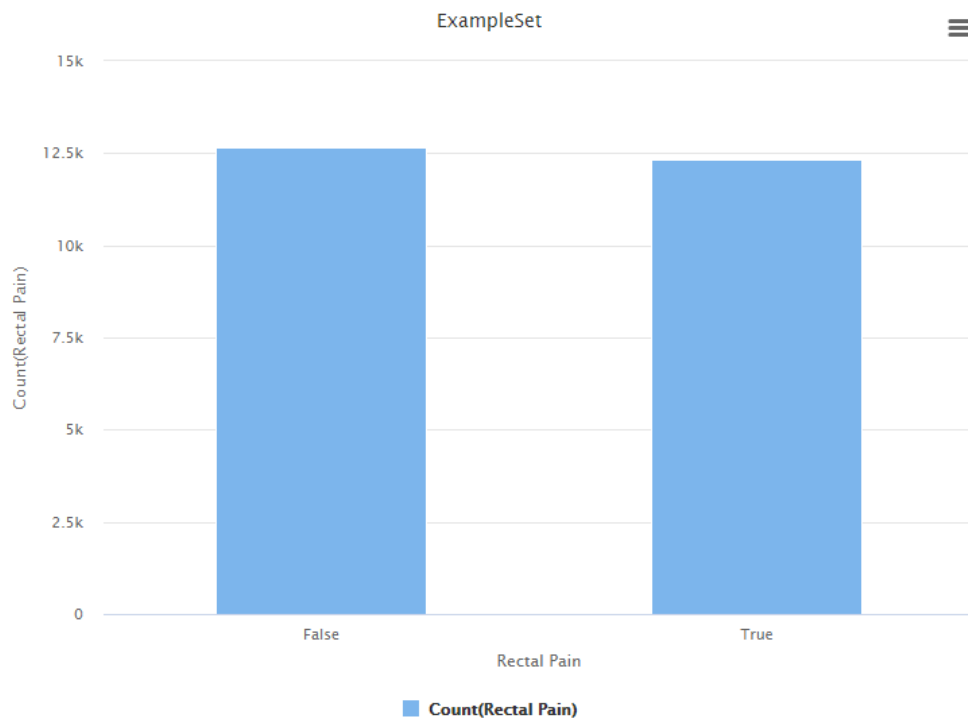
b. Rectal Pain

Rectal Pain merupakan nyeri rectum atau rasa sakit yang muncul di anus. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari Rectal Pain dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 3 Nominal Values pada Rectal Pain

No	Nominal Value	Absolute Count	Fraction
1	False	12655	0.506
2	True	12345	0.494

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Rectal Pain. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 7 Visualisasi Atribut Rectal Pain

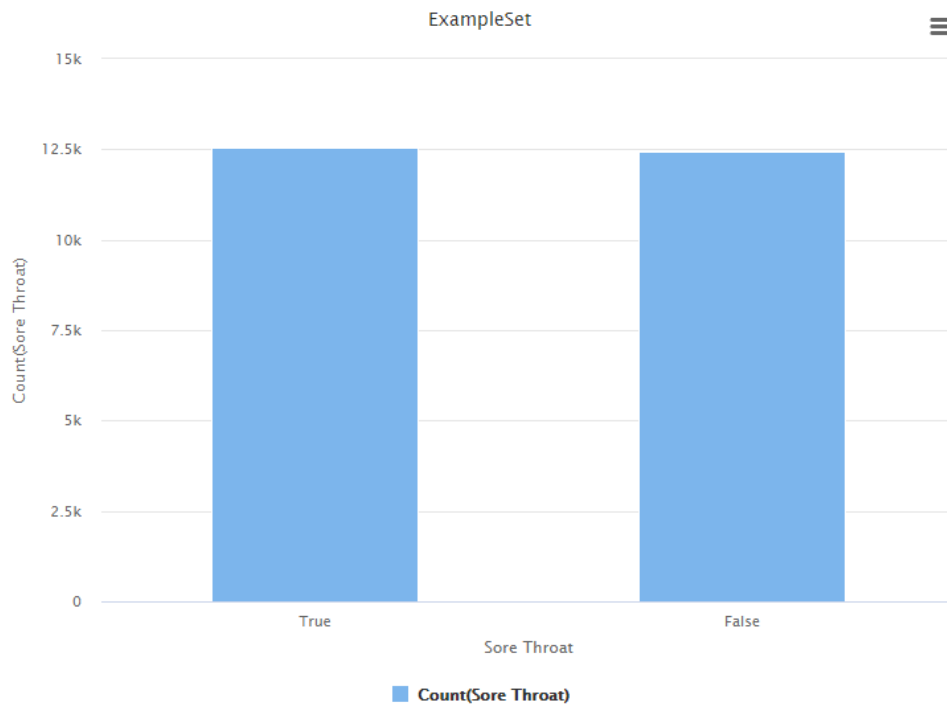
c. Sore Throat

Sore Throat merupakan sakit tenggorokan dimana gejala ini dapat ditemui pada berbagai kondisi medis seperti influenza atau infeksi virus lainnya termasuk virus monkeypox. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari Sore Throat dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 4 Nominal Values pada Sore Throat

No	Nominal Value	Absolute Count	Fraction
1	False	12446	0.498
2	True	12554	0.502

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Sore Throat. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 8 Visualisasi Atribut Sore Throat

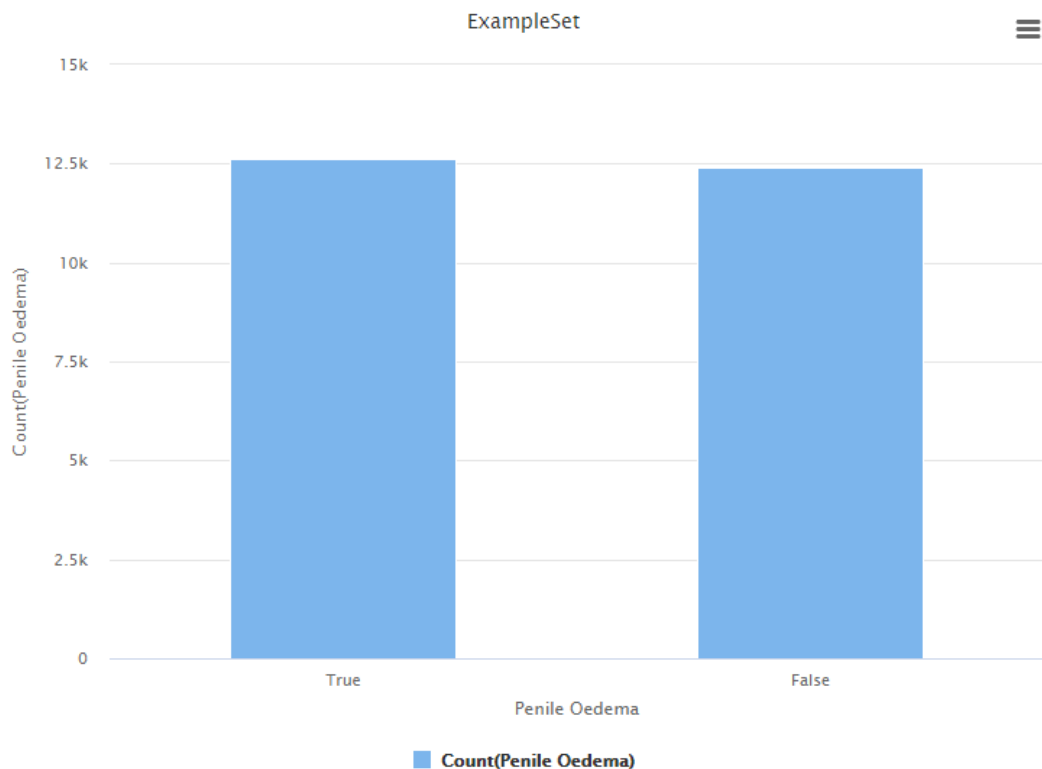
d. Penile Oedema

Penile Oedema atau bisa disebut edema penis merupakan kondisi dimana terjadi pembesaran cairan di ruang intersistital sehingga menimbulkan pembekakan. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari Penile Oedema dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 5 Nominal Values pada Penile Oedema

No	Nominal Value	Absolute Count	Fraction
1	True	12602	0.502
2	False	12388	0.496

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Penile Oedema. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 9 Visualisasi Atribut Penile Oedema

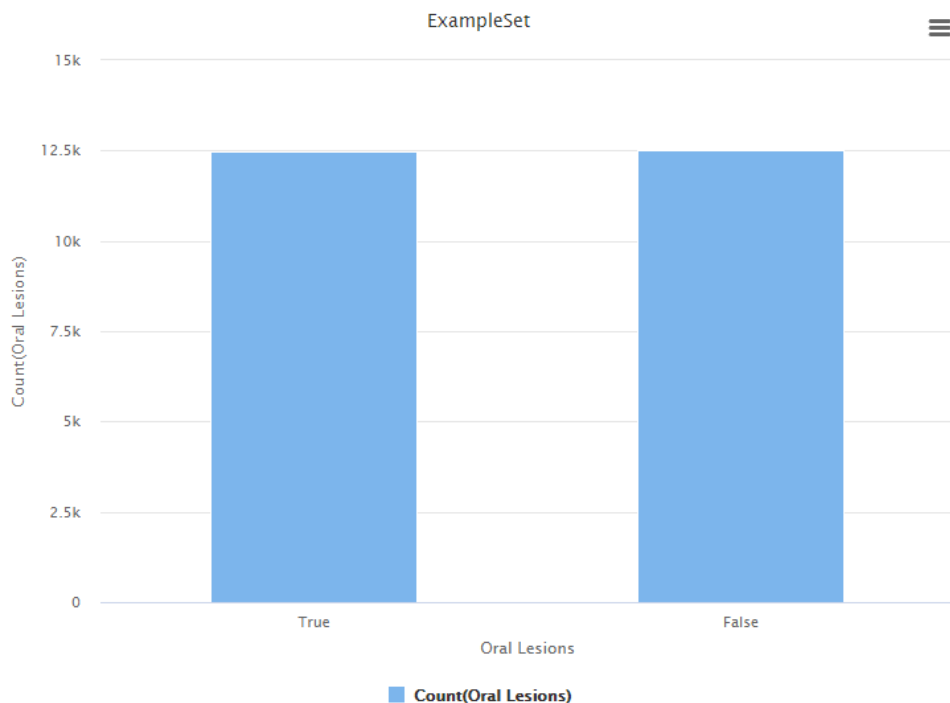
e. Oral Lesions

Oral Lesions atau istilah dalam Indonesia dikenal dengan sariawan merupakan area abnormal dan seringkali menyakitkan yang berkembang di selaput lender. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari Oral Lesions dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 6 Nominal Values pada Oral Lesions

No	Nominal Value	Absolute Count	Fraction
1	False	12514	0.501
2	True	12486	0.499

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Oral Lesions. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 10 Visualisasi Atribut Oral Lesions

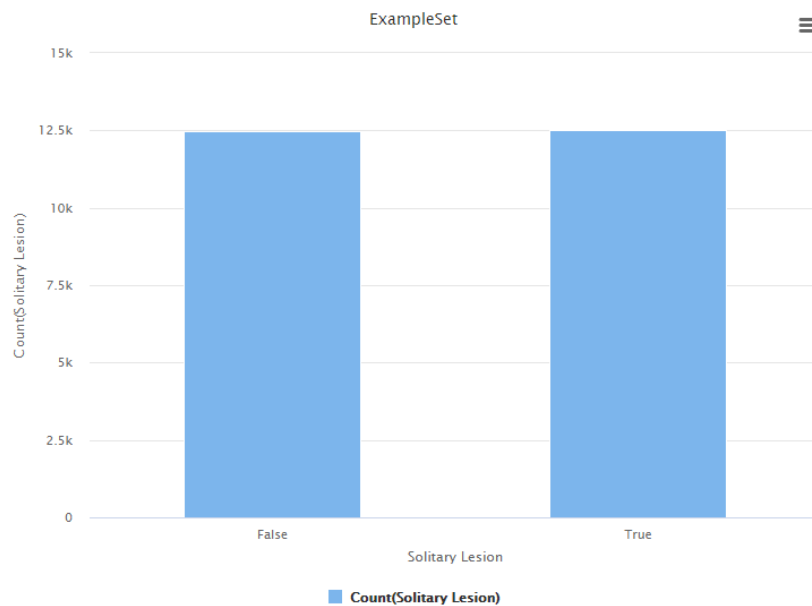
f. Solitary Lesion

Solitary Lesion atau lesi soliter merupakan perubahan patologis yang hanya terjadi dalam satu lokasi atau daerah tertentu dalam tubuh. Dalam konteks medis, lesi mengacu pada area berubah atau rusak dalam jaringan tubuh yang mungkin disebabkan oleh berbagai faktor, seperti infeksi, trauma, inflamasi, atau pertumbuhan abnormal sel. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari Solitary Lesion dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 7 Nominal Values pada Solitary Lesion

No	Nominal Value	Absolute Count	Fraction
1	True	12527	0.501
2	False	12473	0.499

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Solitary Lesion. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 11 Visualisasi Atribut Solitary Lesion

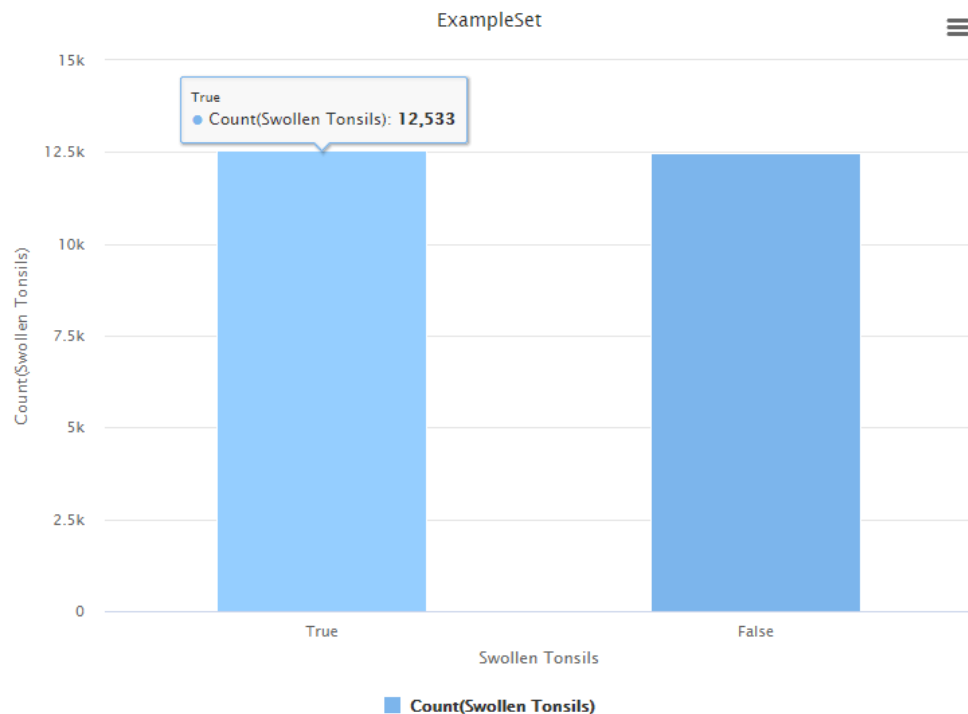
g. Swollen Tonsils

Swollen Tonsils merupakan Radang Amandel Radang amandel atau tonsilitis adalah kondisi ketika amandel mengalami inflamasi atau peradangan. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari Swollen Tonsils dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 8. Nominal Values pada Swollen Tonsils

No	Nominal Value	Absolute Count	Fraction
1	True	12533	0.501
2	False	12467	0.499

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Swollen Tonsils. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 12 Visualisasi Attribut Swollen Tonsils

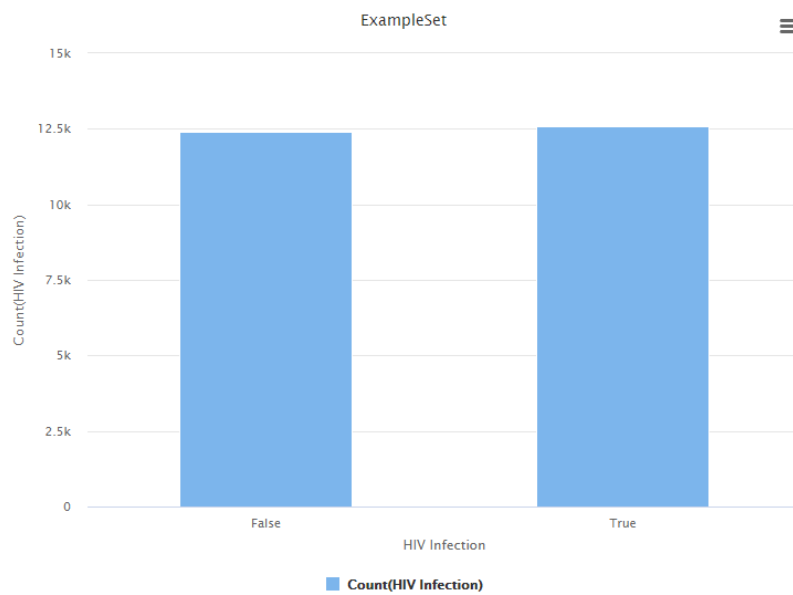
h. HIV Infection

HIV Infection adalah virus yang menyerang sistem kekebalan tubuh yang dapat melemahkan kemampuan tubuh melawan infeksi dan penyakit. HIV menyebabkan AIDS dan mengganggu kemampuan tubuh melawan infeksi. Virus ini dapat ditularkan melalui kontak dengan darah yang terinfeksi, air mani, atau cairan vagina. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari HIV Infection dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 9. Nominal Values pada HIV Infection

No	Nominal Value	Absolute Count	Fraction
1	True	12584	0.503
2	False	12416	0.497

Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam HIV Infection. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 13 Visualisasi Atribut HIV Infection

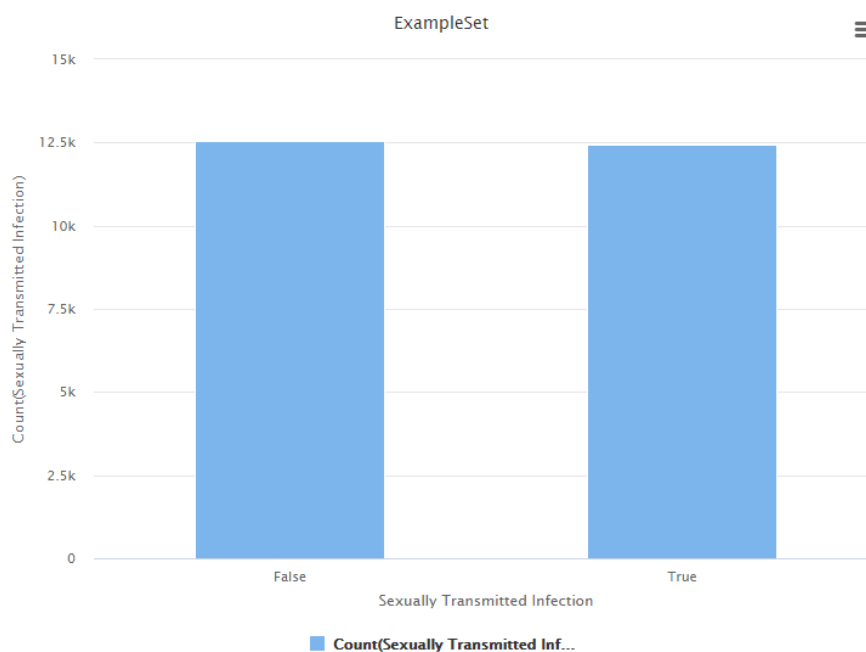
i. Sexually Transmitted Infection

Sexually Transmitted Infection atau STD adalah penyakit menular seksual yang dapat terjadi karena adanya kontak langsung dengan cairan tubuh. Penyakit ini juga di sebabkan karna infeksi. Infeksi yang ditularkan melalui hubungan seksual, disebabkan oleh bakteri, virus, atau parasit. Data yang divisualisasikan pada gambar dibawah ini adalah atribut dari Sexually Transmitted Infection dimana dari keseluruhan data yang terecord adalah 25.000 data dengan data pada masing-masing kriteria sebagai berikut.

Tabel 4. 10. Nominal Values pada Sexually Transmitted Infection

No	Nominal Value	Absolute Count	Fraction
1	True	12554	0.502
2	False	12446	0.498

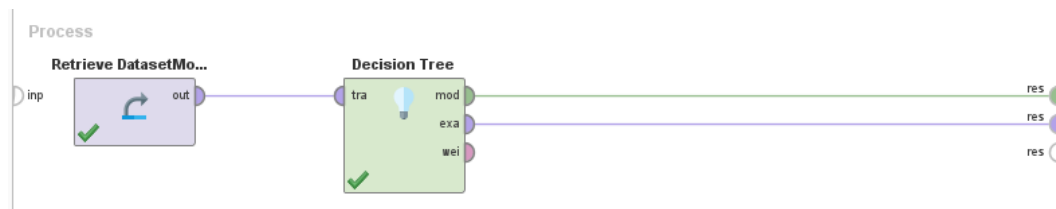
Data diatas adalah jumlah masing-masing data pada setiap kriteria di dalam Sexually Transmitted Infection. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 14 Visualisasi Atribut Sexually Transmitted Infection

4.1.3. Pemrosesan Algoritma Decision Tree C4.5

Pada tahap pemrosesan ini adalah pemrosesan pada algoritma di mana dataset yang sudah di import dilakukan klasifikasi dengan menggunakan model algoritma Decision Tree C4.5 agar dapat menemukan pola dari pohon keputusan dengan menentukan atribut manakah yang memiliki gain ratio tertinggi sehingga menjadi induk dari pohon keputusan dalam menentukan prediksi akurasi pada penyakit monkeypox tersebut berdasarkan pola dari pohon keputusan.



Gambar 4. 15 Pemrosesan Decision Tree C4.5

Pada tampilan diatas merupakan pemrosesan pada algoritma Decision tree C4.5. Setelah pemrosesan dijalankan maka akan menghasilkan tampilan pohon keputusan pada algoritma Decection Tree C4.5 diaman node pada atribut awal yaitu Systemic Illness.



Gambar 4. 16 Pohon Keputusan Decision Tree C4.5

Dapat dilihat bahwa pada tampilan pohon keputusan diatas menampilkan attribute Systemic Illness pada awal dari pohon keputusan dimana dalam menentukan seseorang terinfeksi positive maupun negative di tentukan dalam atribut Systemic

Illness dengan keterangan fever dan Swollen Lymph Node untuk terinfeksi positive pada penyakit Monkeypox dan menurun untuk seterusnya sesuai pada pohon keputusan dalam Decission Tree DTC4.5. Berikut adalah deskripsi pada pohon keputusan diatas.

Tree

```

Systemic Illness = Fever: Positive {Negative=1602, Positive=4780}
Systemic Illness = Muscle Aches and Pain
| Rectal Pain = False
| | HIV Infection = False: Negative {Negative=1056, Positive=517}
| | HIV Infection = True: Positive {Negative=726, Positive=861}
| Rectal Pain = True
| | Sexually Transmitted Infection = False
| | | HIV Infection = False
| | | | Sore Throat = False
| | | | | Swollen Tonsils = False
| | | | | Penile Oedema = False: Negative {Negative=67, Positive=18}
| | | | | Penile Oedema = True
| | | | | | Solitary Lesion = False: Negative {Negative=27, Positive=9}
| | | | | | Solitary Lesion = True
| | | | | | Oral Lesions = False: Negative {Negative=17, Positive=8}
| | | | | | Oral Lesions = True: Positive {Negative=9, Positive=11}
| | | | | Swollen Tonsils = True: Negative {Negative=99, Positive=76}
| | | | Sore Throat = True
| | | | | Penile Oedema = False
| | | | | Solitary Lesion = False
| | | | | | Swollen Tonsils = False: Positive {Negative=21, Positive=24}
| | | | | | Swollen Tonsils = True: Negative {Negative=39, Positive=15}
| | | | | Solitary Lesion = True
| | | | | | Swollen Tonsils = False: Negative {Negative=23, Positive=17}
| | | | | | Swollen Tonsils = True: Positive {Negative=17, Positive=27}
| | | | Penile Oedema = True
| | | | | Swollen Tonsils = False: Positive {Negative=47, Positive=55}
| | | | Swollen Tonsils = True
| | | | | Swollen Tonsils = True
| | | | | Oral Lesions = False: Positive {Negative=12, Positive=34}
| | | | | Oral Lesions = True
| | | | | | Solitary Lesion = False: Negative {Negative=9, Positive=8}
| | | | | | Solitary Lesion = True: Positive {Negative=10, Positive=23}
| | | HIV Infection = True: Positive {Negative=302, Positive=478}
| | Sexually Transmitted Infection = True: Positive {Negative=445, Positive=1043}
Systemic Illness = None
| Rectal Pain = False
| | HIV Infection = False
| | | Sexually Transmitted Infection = False
| | | | Penile Oedema = False: Negative {Negative=318, Positive=51}
| | | | Penile Oedema = True
| | | | | Sore Throat = False: Negative {Negative=150, Positive=40}
| | | | | Sore Throat = True
| | | | | Oral Lesions = False: Negative {Negative=71, Positive=29}
| | | | | Oral Lesions = True
| | | | | | Solitary Lesion = False: Positive {Negative=25, Positive=26}
| | | | | | Solitary Lesion = True: Negative {Negative=38, Positive=20}
| | | Sexually Transmitted Infection = True
| | | | Oral Lesions = False
| | | | | Penile Oedema = False: Negative {Negative=139, Positive=60}
| | | | | Penile Oedema = True
| | | | | Swollen Tonsils = False: Negative {Negative=62, Positive=36}
| | | | | Swollen Tonsils = True
| | | | | | Solitary Lesion = False: Negative {Negative=31, Positive=17}
| | | | | | Solitary Lesion = True: Positive {Negative=22, Positive=34}
| | | | Oral Lesions = True: Negative {Negative=210, Positive=210}
| | HIV Infection = True
| | | Sexually Transmitted Infection = False: Negative {Negative=416, Positive=379}
| | | Sexually Transmitted Infection = True: Positive {Negative=281, Positive=476}

```

```

| Rectal Pain = True
| | HIV Infection = False
| | | Sexually Transmitted Infection = False
| | | | Sore Throat = False
| | | | | Penile Oedema = False: Negative {Negative=126, Positive=76}
| | | | | Penile Oedema = True: Positive {Negative=82, Positive=91}
| | | | Sore Throat = True: Positive {Negative=172, Positive=222}
| | | Sexually Transmitted Infection = True: Positive {Negative=284, Positive=484}
| | HIV Infection = True: Positive {Negative=465, Positive=1073}
Systemic Illness = Swollen Lymph Nodes: Positive {Negative=1671, Positive=4581}

```

Gambar 4. 17 Deskripsi Pohon Keputusan

Pada tampilan hasil diatas dengan menggunakan tool Rapidminer diketahui bahwa node root awal adalah Systemic Illness. Dalam hasil penelitian ini dilakukan pengukuran nilai entropy dan juga gain dengan perhitungan manual menggunakan excel untuk melihat apakah perhitungan menggunakan rapidminer berjalan dengan baik sesuai dengan perhitungan excel. Pada tahap awal akan dilakukan perhitungan nilai total entropy pada dataset.

Tabel 4. 11 Total Nilai Entropy Dataset

	Jumlah Kasus	Monkeypox (Positive)	Monkeypox (Negative)	Entropy
Total	25000	15909	9091	0,94566

$$H(X) = - \sum_{i=1}^n P(xi) \log_2 P(xi)$$

$H(X)$ adalah nilai entropy dari himpunan X dan n adalah jumlah kemungkinan hasil yang berbeda dalam himpunan dihitung. Kemudian $P(xi)$ adalah probabilitas dari hasil xi dalam himpunan data. Probabilitas ini dihitung dengan menghitung jumlah kemunculan xi dibagi dengan total jumlah data. Berikut adalah perhitungan pada nilai entropy dari total dataset.

$$= - \left(\left(\frac{15909}{25000} \right) \times \log_2 \left(\frac{15909}{25000} \right) \right) - \left(\left(\frac{9091}{25000} \right) \times \log_2 \left(\frac{9091}{25000} \right) \right) = 0,94566$$

Pada hasil perhitungan diatas adalah nilai entropy dari total data dalam dataset. Selanjutnya adalah perhitungan pada nilai entropy pada masing-masing Atribut dan nilai dari gainnya.

Tabel 4. 12 Nilai Gain Pada Atribut

No	Atribut	Keterangan	Jumlah Kasus	Monkeypox (Positive)	Monkeypox (Negative)	Entropy	Info Gain
1		Total	25000	15909	9091	0,94566	
2	Systemic Illness	Fever	6382	4780	1602	0,81289	0,03534
		Swollen Lymph Nodes	6252	4581	1671	0,83753	
		Muscle Aches and Pain	6150	3224	2926	0,99831	
		None	6216	3324	2892	0,99651	

Pada table diatas adalah hasil dari entropy dari masing-masing jenis gejala dalam systemic illness dan juga gain info dari systemic illness. Berikut adalah perhitungan manual dari hasil entropy dan gain info dari attribute Systemic Illness berdasarkan beberapa gejala sebagai berikut.

a. *Entropy Fever*

$$= (0,94566) - \left(\left(\frac{4780}{6382} \right) \times \log_2 \left(\frac{4780}{6382} \right) \right) + \left(\left(\frac{1602}{6382} \right) \times \log_2 \left(\frac{1602}{6382} \right) \right)$$

$$= 0,81289$$

Nilai entropy dari gejala Fever adalah 0,81289

b. *Swollen Lymph Nodes*

$$= (0,94566) - \left(\left(\frac{4581}{6252} \right) \times \log_2 \left(\frac{4581}{6252} \right) \right) + \left(\left(\frac{1671}{6252} \right) \times \log_2 \left(\frac{1671}{6252} \right) \right) = 0,83753$$

Nilai entropy dari gejala *Swollen Lymph Nodes* adalah 0,83753

c. *Muscle Aches and Pain*

$$= (0,94566) - \left(\left(\frac{3224}{6150} \right) \times \log_2 \left(\frac{3224}{6150} \right) \right) + \left(\left(\frac{2926}{6150} \right) \times \log_2 \left(\frac{2926}{6150} \right) \right) = 0,99831$$

Nilai entropy dari gejala *Muscle Aches and Pain* adalah 0,99831

d. *None*

$$= (0,94566) - \left(\left(\frac{3324}{6216} \right) \times \log_2 \left(\frac{3324}{6216} \right) \right) + \left(\left(\frac{2892}{6216} \right) \times \log_2 \left(\frac{2892}{6216} \right) \right) = 0,99651$$

Nilai entropy dari systemic illness yang tidak memiliki gejala atau None adalah 0,99651

Perhitungan manual diatas merupakan perhitungan manual dari Systemic illness dari beberapa gejala yang ada di dalamnya.

e. Info Gain

Setelah di tentukan nilai entropy dari masing-masing gejala yang ada didalam Systemic Illanes maka dilakukan perhitungan info gain sesuai pada table 4.12. berikut adalah perhitungan manual dari info gain.

$$\begin{aligned} &= (0,94566) - \left(\left(\frac{6382}{25000} \right) \times 0,81286 \right) - \left(\left(\frac{6252}{25000} \right) \times 0,83753 \right) \\ &\quad - \left(\left(\frac{6150}{25000} \right) \times 0,99831 \right) - \left(\left(\frac{6216}{25000} \right) \times 0,99651 \right) \\ &= 0,03534 \end{aligned}$$

Dari perhitungan nilai pada info gain diatas adalah 0,03534

f. Split Info

Setelah perhitungan info gain selanjutnya dilakukan perhitungan split info dimana dalam perhitungan ini dilakukan agar nantinya dapat menentukan nilai dari gain ratio dengan membagi antara info gain dan split info. Berikut adalah perhitunganya berdasarkan attribute Systemic Illness.

$$\begin{aligned} &= \left(- \left(\frac{6382}{25000} \right) \times \log_2 \left(\frac{6382}{25000} \right) \right) - \left(\left(\frac{6252}{25000} \right) \times \log_2 \left(\frac{6252}{25000} \right) \right) - \\ &\left(- \left(\frac{6150}{25000} \right) \times \log_2 \left(\frac{6150}{25000} \right) \right) - \left(- \left(\frac{6216}{25000} \right) \times \log_2 \left(\frac{6216}{25000} \right) \right) = 1,99987 \end{aligned}$$

Didapatkan nilai pada perhitungan split info sebesar 1,99987

g. Gain Ratio

Setelah perhitungan entropy, info gain dan split info selesai dilakukan maka dilakukan perhitungan Gain Ratio dimana pada perhitungan ini membagi hasil dari perhitungan Info Gain dan Split Info.

$$= \frac{0,03534}{1,99987} = 0,01767$$

Maka hasil dari perhitungan Gain Ratio dalam attribute Systemic Illness adalah 0,01767. Pada attribut selanjutnya dilakukan perhitungan yang sama dengan mengikuti proses perhitungan yang dilakukan pada attribute Systemic Illness. Dan berikut adalah hasil dari seluruh attribute yang dilakukan dengan perhitungan manual.

Tabel 4. 13 Hasil Perhitungan Manual Seluruh Attribut

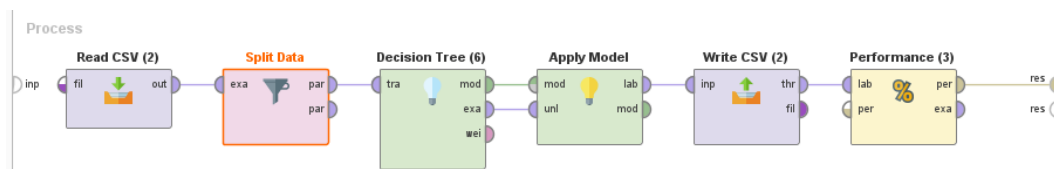
Attribut	Keterangan	Jumlah Kasus	Monkeypox (Negative)		Entropy	Info Gain	Split Info	Gain Ratio
			Positive	Negative				
	Total	25000	15909	9091	0,94566			
Systemic Illness						0,03534	1,99987	0,01767
	Fever	6382	4780	1602	0,81289			
	Swollen Lymph Nodes	6252	4581	1671	0,83753			
	Muscle Aches and Pain	6150	3224	2926	0,99831			
	None	6216	3324	2892	0,99651			
Rectal Pain	FALSE	12655	7207	5448	0,98602	0,01436	0,99989	0,01436
	TRUE	12345	8702	3643	0,87522			
Sore Throat	FALSE	12446	7539	4907	0,9675	0,0029	0,99999	0,0029
	TRUE	12554	8370	4184	0,91824			
Penile Oedema	FALSE	12388	7511	4877	0,96714	0,00277	0,99994	0,00277
	TRUE	12612	8398	4214	0,91909			
	FALSE	12514	7651	4863	0,96389	0,00195	1	0,00195

Oral Lesions	TRUE	12486	8258	4228	0,92349			
Solitary Lesion	FALSE	12473	7712	4761	0,95924	0,00101	1	0,00101
	TRUE	12527	8197	4330	0,93013			
Swollen Tonsils	FALSE	12467	7854	4613	0,95069	0,00013	0,99999	0,00013
	TRUE	12533	8055	4478	0,94042			
HIV Infection	FALSE	12416	7022	5394	0,98756	0,01548	0,99997	0,01548
	TRUE	12584	8887	3697	0,87356			
Sexually Transmitted Infection	FALSE	12554	7247	5307	0,9827	0,01102	0,99999	0,01102
	TRUE	12446	8662	3784	0,88617			

Dari perhitungan manual diatas dapat dilihat bahwa nilai tertinggi gain ratio adalah pada attribute Systemic Illness dengan nilai 0,01767. Hal ini sesuai dengan pengolahan decision tree yang dilakukan dalam tool rapidminer dimana Systemic Illness adalah node root awal setelah dilakukan pemrosesan.

4.1.4. Pemrosesan Decision Tree C4.5 dengan Split Validation

Pada tahapan split validation digunakan untuk membagi data menjadi dua bagian secara acak. Sebagaimana sebagai data training dan sebagian sebagai data testing. Pada pembagian split validation di dalam penelitian ini dilakukan dengan pembagian 80:20 dalam split validation. Berikut adalah alur pemrograman pada tool rapidminer dengan split validation.



Gambar 4. 18 Pemrosesan DTC4.5 dengan Split Validation

Dalam alur pemrograman dengan rapidminer diatas menggunakan seluruh dataset yang digunakan yaitu berjumlah 25000 dengan 10 atribut yang digunakan. Pada split validation digunakan perbandingan 80:20 untuk membagi dataset secara acak

dengan menggunakan decision tree untuk melakukan prediksi pada infeksi wabah monkeypox menggunakan parameter positive dan negative. Kemudian digunakan apply model untuk menerapkan data yang telah dilatih pada pemrosesan decision tree. Setelah itu digunakan operator write csv yang bertujuan menyimpan data yang telah dilatih untuk digunakan sebagai validasi hasil. Pada tahapan ini juga digunakan operator performance untuk melihat nilai akurasi pada data yang dilatih. Dan berikut adalah hasilnya.

PerformanceVector

```

PerformanceVector:
accuracy: 69.43%
ConfusionMatrix:
True:  Negative      Positive
Negative:      2455      1296
Positive:      4818      11431
precision: 70.35% (positive class: Positive)
ConfusionMatrix:
True:  Negative      Positive
Negative:      2455      1296
Positive:      4818      11431
recall: 89.82% (positive class: Positive)
ConfusionMatrix:
True:  Negative      Positive
Negative:      2455      1296
Positive:      4818      11431
AUC (optimistic): 0.720 (positive class: Positive)
AUC: 0.677 (positive class: Positive)
AUC (pessimistic): 0.635 (positive class: Positive)

```

Gambar 4. 19 Performance Vector

Pada performance vector dapat dilihat bahwa hasil akurasi masih relative sangat kecil yaitu sebesar 69% dan hasil presisi sebesar 70 % namun hasil yang cukup tinggi di dapatkan yaitu pada recall yaitu sebesar 89% sehingga masih perlu pengolahan agar nilai akurasi dalam prediksi wabah penyakit monkeypox mendapatkan hasil yang tinggi. Kemudian pada hasil example Set yang didapatkan adalah sebagai berikut.

prediction(MonkeyPox)	Systemic Ill...	Rectal Pain	Sore Throat	Penile Oede...	Oral Lesions	Solitary Lesi...	Swollen Ton...	HIV Infection	MonkeyPox	confidence(Ne...
Negative	None	False	True	True	True	False	True	False	Negative	0.668
Positive	Fever	True	False	True	True	False	False	True	Positive	0.253
Positive	Fever	False	True	True	False	False	False	True	Positive	0.253
Positive	None	True	False	False	False	True	True	True	Positive	0.371
Positive	Swollen Lym...	True	True	True	False	False	True	True	Positive	0.264
Positive	Swollen Lym...	False	True	False	False	False	False	False	Negative	0.264
Positive	Fever	False	True	False	False	False	False	True	Positive	0.253
Positive	Fever	True	True	False	True	True	True	False	Positive	0.253
Negative	Muscle Ache...	False	True	True	True	False	False	False	Positive	0.603
Positive	Fever	False	False	True	True	True	False	True	Negative	0.253
Negative	Muscle Ache...	False	True	True	True	True	True	False	Negative	0.603
Positive	Fever	True	False	True	False	True	True	True	Positive	0.253
Positive	Swollen Lym...	True	True	False	True	True	True	False	Positive	0.264

Gambar 4. 20 Tampilan Hasil Example Set

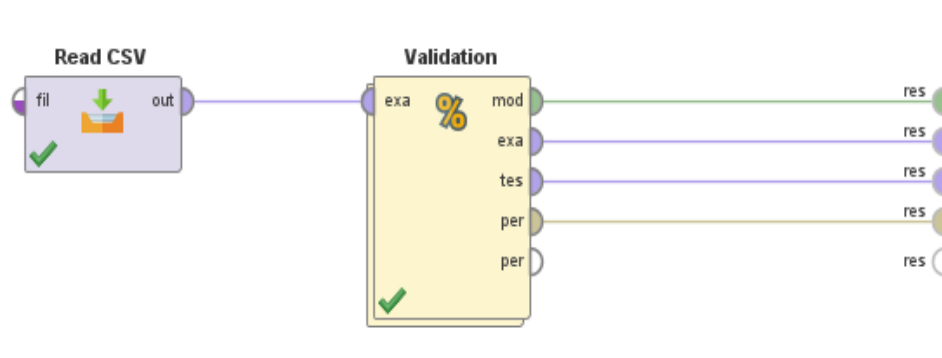
Pada example set dapat dilihat bahwa label monkeypox berhasil di prediksi dalam pemodelan untuk meningkatkan akurasi pada prediksi wabah monkeypox. Nantinya hasil dari prediksi dari label atribut monkeypox ini akan digunakan sebagai label dalam prediksi monkeypox dalam tahapan validasi.

4.1.5. Validasi Hasil Pemrosesan Algoritma

Pada validasi hasil pemrosesan algoritma adalah tahapan ahir dimana akan digunakan cross validation dengan k-fold 10 untuk melihat tingkat akurasi pada prediksi wabah monkeypox dengan decision tree C4.5 dan Particle Swarm Optimatation.

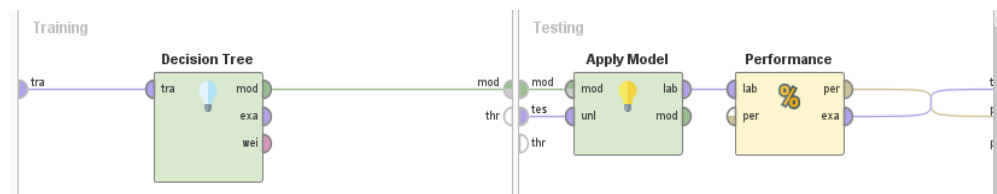
a. Validasi Hasil *Decision Tree* C4.5 dengan *Cross Validation*

Pada validasi dengan Cross Validation adalah untuk memperoleh hasil akurasi yang maksimal dalam melakukan prediksi pada wabah monkeypox. Adapun alur dalam pemrosesan decision tree C4.5 dengan Cross Validation adalah sebagai berikut.



Gambar 4. 21 Cross Validation dengan DTC4.5

Pada pemrosesan diatas adalah alur dari cross validation yang mana didalamnya terdapat sub untuk menjalankan algoritma yang akan di jalankan. Algoritma yang ada didalam cross validation adalah decision tree. Berikut adalah tampilan alur pemrosesanya.



Gambar 4. 22 Sub alur Pemrosesan Cross Validation

Pada sub alur cross validation terdapat decision tree sebagai algoritma untuk pengolahan pada data training dan operator apply data untuk testing data kemudian pada operator performance digunakan untuk melihat hasil dari pemrosesan dengan rapid miner tersebut. Dan berikut adalah hasilnya.

accuracy: 81.24% +/- 0.77% (micro average: 81.25%)

	true Negative	true Positive	class precision
pred. Negative	0	0	0.00%
pred. Positive	3751	16249	81.25%
class recall	0.00%	100.00%	

Gambar 4. 23 Hasil Accuracy

Pada hasil akurasi dari pemrosesan algoritma Decision Tree dengan Cross Validation adalah sebesar 81%

precision: 81.24% +/- 0.77% (micro average: 81.25%) (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	0	0	0.00%
pred. Positive	3751	16249	81.25%
class recall	0.00%	100.00%	

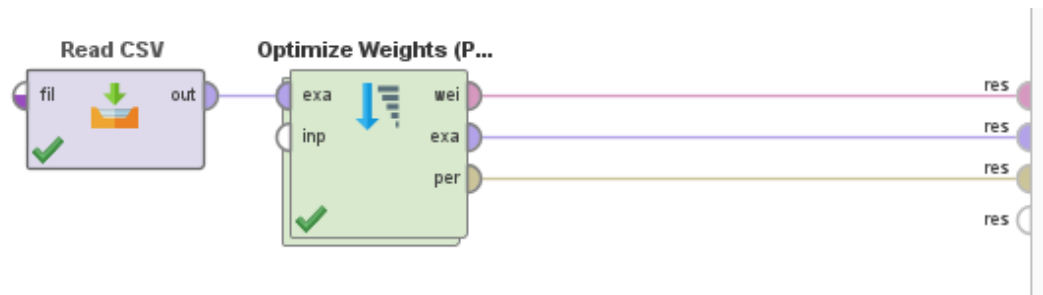
Gambar 4. 24 Hasil Precision

Pada hasil presisi dari pemrosesan algoritma Decision Tree dengan Cross Validation adalah sebesar 81,24 %

Pada pemrosesan akurasi dan presisi masih terbilang belum cukup tinggi karna masih menggunakan satu model saja.

b. Validasi Hasil *Partcle Swarm Optimization* dengan *Decision Tree C4.5*

Pada validasi dengan *Partcle Swarm Optimization* dan *Decision Tree C4.5* adalah untuk memperoleh hasil akurasi yang maksimal dalam melakukan prediksi. Penggunaan PSO untuk melakukan optimasi pada pengolahan algoritma decision tree agar hasil akurasi lebih maksimal dan mendapat nilai akurasi yang tinggi. Pada tahapan ini juga digunakan Cross validation untuk memaksimalkan pemrosesan agar mendapat hasil yang baik. Adapun alur dalam pemrosesan *Partcle Swarm Optimization* dan *Decision Tree C4.5* adalah sebagai berikut.



Gambar 4. 25 Partcle Swarm Optimization dengan Decision Tree C4.5

Pada pemrosesan *Partcle Swarm Optimization* dengan *Decision Tree C4.5* adalah tahapan ahir untuk melihat apakah nilai akurasi mejadi lebih baik atau tidak. Pada pemrosesan *Partcle Swarm Optimization* terdapat sub proses

dimana didalam sub proses akan dijalankan operator cross validation seperti gambar berikut.



Gambar 4. 26 Sub Proses Partcle Swarm Optimization (PSO)

Setelah memasukan operator Cross Validation masih terdapat sub alur di dalam cross validation yang dapat dilihat persis seperti gambar 4.22. kemudian setelah pemrosesan rapidminer dijalankan maka berikut adalah hasil dari optimasi *Partcle Swarm Optimization* dengan *Decision Tree C4.5*.

accuracy: 99.51% +/- 0.09% (micro average: 99.51%)

	true Negative	true Positive	class precision
pred. Negative	3712	59	98.44%
pred. Positive	39	16190	99.76%
class recall	98.96%	99.64%	

Gambar 4. 27 Hasil Accuracy Partcle Swarm Optimization dengan Decision Tree C4.5

Pada hasil akurasi dari optimasi *Partcle Swarm Optimization* dengan *Decision Tree C4.5* dan juga Cross Validation adalah sebesar 99,51%. Kemudian untuk hasil Presisinya adalah sebagai berikut.

precision: 99.76% +/- 0.16% (micro average: 99.76%) (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	3712	59	98.44%
pred. Positive	39	16190	99.76%
class recall	98.96%	99.64%	

Gambar 4. 28. Hasil Presisi Partcle Swarm Optimization dengan Decision Tree C4.5

Pada hasil Presisi dari optimasi *Partcle Swarm Optimization* dengan *Decision Tree C4.5* dan juga Cross Validation adalah sebesar 99,76%. Kemudian untuk hasil recallnya adalah sebagai berikut

recall: 99.64% +/- 0.10% (micro average: 99.64%) (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	3712	59	98.44%
pred. Positive	39	16190	99.76%
class recall	98.96%	99.64%	

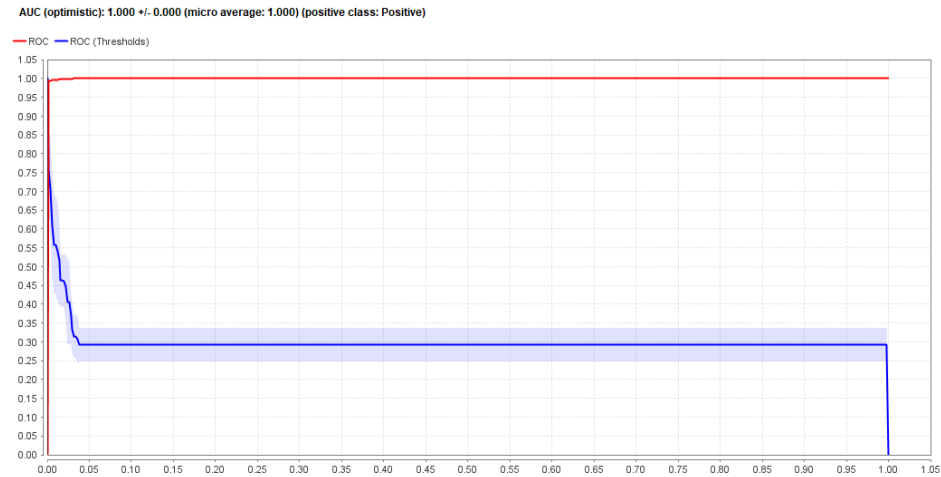
Gambar 4. 29 Hasil Recall Partcle Swarm Optimation dengan Decision Tree C4.5

Pada hasil Recall dari optimasi *Partcle Swarm Optimization* dengan *Decision Tree C4.5* dan juga Cross Validation adalah sebesar 99,64%. Pada hasil dari ketiga akurasi ini mendapatkan peningkatan yang lebih baik dari pemrosesan dengan decision tree dan cross validation sebelumnya. Pada Attribute Weights dalam optimasi Weight (PSO) adalah sebagai berikut.

attribute	weight
Patient_ID	0.267
Systemic Illness	0.953
Rectal Pain	0.043
Sore Throat	0.719
Penile Oedema	0.637
Oral Lesions	0
Solitary Lesion	0.002
Swollen Tonsils	0.539
HIV Infection	0.081
Sexually Transmitted Infection	1
MonkeyPox	0.081
confidence(Negative)	0
confidence(Positive)	0

Gambar 4. 30 Attribute Weight

Setelah melihat hasil akurasi yang sangat baik dari nilai akurasi, presisi dan recall. Berikut adalah nilai dari ROC AUC dari hasil pemrosesan *Particle Swarm Optimization* dengan *Decision Tree C4.5*.



Gambar 4. 31. ROC AUC

Pada hasil ROC AUC mendapat nilai 1.000+/- 0.000. Pada garis kurva ROC AUC dapat dilihat bahwa nilai ROC yaitu pada garis merah mendapatkan hasil kurva yang sempurna pada nilai 1.000 sedangkan pada nilai ROC Thresholds atau pada garis kurva warna biru mendapatkan nilai yang relative stabil namun tidak pada garis ahir dengan kondisi Thresholds yang stabil di angka 0.30. Berikut adalah table ulasan hasil dari pemrosesan *Particle Swarm Optimization* dengan *Decision Tree C4.5*.

Tabel 4. 14. Hasil dari PSO dan Decision Tree dengan Cross Validation

No	PSO + DTC4.5	Hasil
1	Accuracy	99,51%.
2	Precision	99,76%.
3	Recall	99,64%.
4	ROC AUC	1.000

4.2. Pembahasan Hasil Penelitian

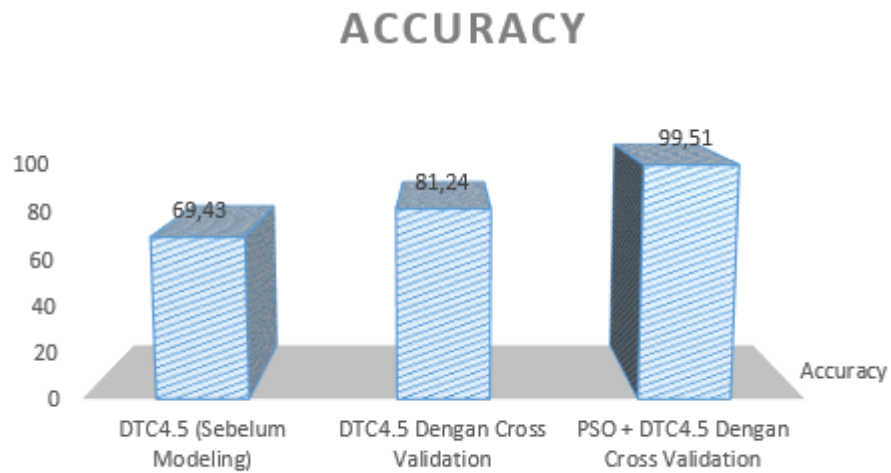
Pembahasan pada hasil penelitian membahas tentang penelitian yang telah dilakukan. Pada penelitian ini mengklasifikasi wabah penyakit Monkeypox atau

biasa di sebut cacar monyet dengan menggunakan *Particle Swarm Optimization* dan *Decision Tree C4.5*. pada penelitian ini melakukan optimasi pada algoritma *Decision Tree C4.5* dengan menggunakan *Particle Swarm Optimization* sebagai algoritma untuk melakukan optimasi dalam meningkatkan akurasi. Pada penelitian ini menggunakan dataset public yang kemudian dilakukan split validation dengan perbandingan 80:20. Pada pemrosesan awal pada saat pemodelan ketika melihat hasil akurasi dengan performance vector terbilang mendapat hasil yang sangat kecil namun setelah dilakukan tahapan modeling untuk memprediksi parameter Positive dan negative dari Atribut label yaitu Monkeypox mendapatkan peningkatan akurasi yang signifikan. Berikut adalah perbandingan hasil sebelum dan sesudah tahapan optimasi.

Tabel 4. 15. Hasil dari PSO dan Decision Tree dengan Cross Validation

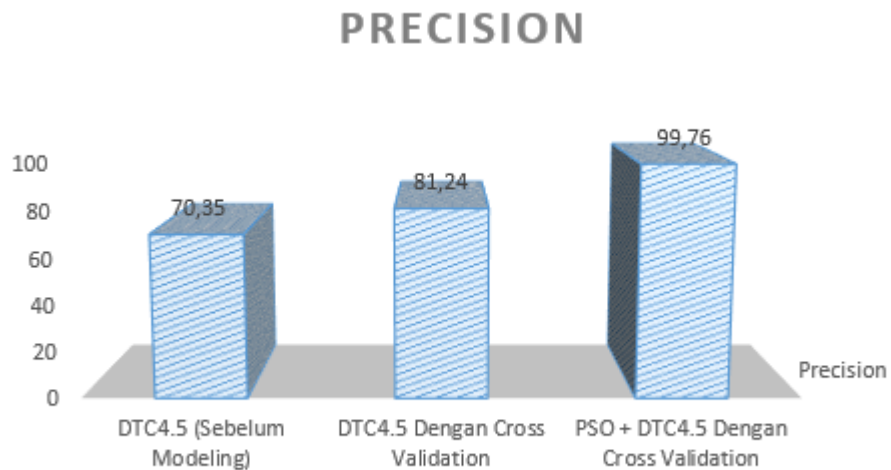
No	Keterangan	DTC4.5 (Sebelum Modeling)	DTC4.5 Dengan Cross Validation	PSO + DTC4.5 Dengan Cross Validation
1	Accuracy	69,43 %	81, 24 %	99,51%.
2	Precision	70,35 %	81,24 %	99,76%.
3	Recall	89,82 %	100 %	99,64%.
4	ROC AUC	0,720	1.000	1.000

Pada table perbandingan dalam pembahasan hasil diatas dapat dilihat hasil yang cukup signifikan dalam kenaikan tingkat akurasi yang pada awalnya hanya 69,43 % hingga mencapai 99,51 %. Dapat disimpulkan bahwa pemodelan dengan melakukan prediksi label dan juga optimasi algoritma Decision Tree dengan PSO dapat meningkatkan akurasi dengan sangat baik. Berikut adalah grafik perbandingan dan sekaligus kenaikan nilai darisebelum modeling sampai dengan optimasi.



Gambar 4. 32 Grafik Perbandingan Accuracy

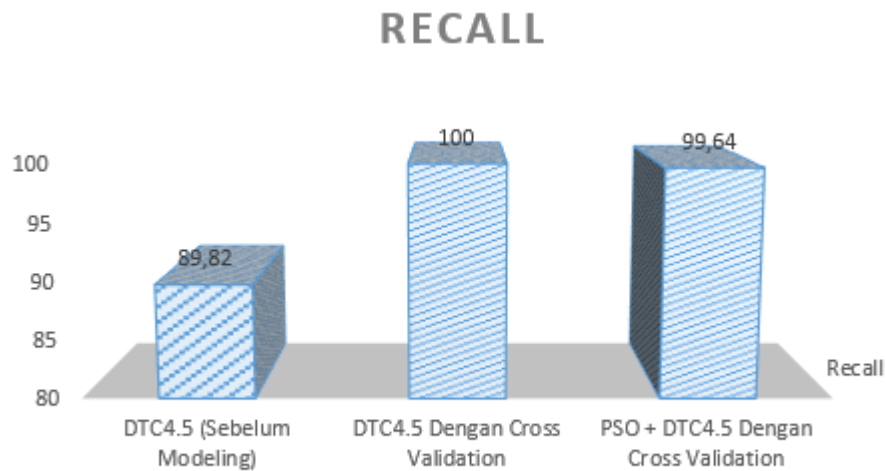
Grafik diatas merupakan grafik perbandingan Accuracy dari ketiga pemrosesan yang dilakukan dalam penelitian ini dimana nilai tertinggi dari akurasi adalah pemrosesan pada algoritma PSO + DTC4.5 Dengan Cross Validation dengan nilai 99,51%.



Gambar 4. 33. Grafik Perbandingan Precision

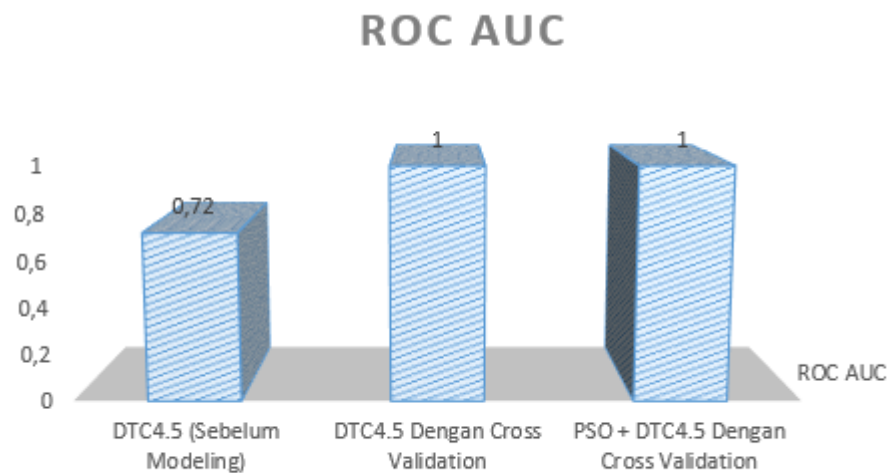
Grafik diatas merupakan grafik perbandingan Precision dari ketiga pemrosesan yang dilakukan dalam penelitian ini dimana nilai tertinggi dari presisi adalah

pemrosesan pada algoritma PSO + DTC4.5 Dengan Cross Validation dengan nilai 99,76%.



Gambar 4. 34. Grafik Perbandingan Recall

Grafik diatas merupakan grafik perbandingan Recall dari ketiga pemrosesan yang dilakukan dalam penelitian ini dimana nilai tertinggi dari recall adalah pemrosesan pada algoritma PSO + DTC4.5 Dengan Cross Validation dengan nilai 99,64%.



Gambar 4. 35. Grafik Perbandingan ROC AUC

Grafik diatas merupakan grafik perbandingan ROC AUC dari ketiga pemrosesan yang dilakukan dalam penelitian ini dimana nilai tertinggi dari recall terdapat pada

dua pemrosesan yaitu pemrosesan pada DTC4.5 dengan Cross Validation dan PSO + DTC4.5 Dengan Cross Validation mendapatkan nilai yang sama yaitu 1,000. Pada Garis kurva yang dapat dilihat pada gambar 4.31 dapat di Analisa bahwa nilai ROC yaitu pada garis merah mendapatkan hasil kurva yang sempurna pada nilai 1.000 sedangkan pada nilai ROC Thresholds atau pada garis kurva warna biru mendapatkan nilai yang relative stabil namun tidak pada garis ahir dengan kondisi Thresholds yang stabil di angka 0.30.