



LAMPIRAN-LAMPIRAN

Lampiran 1. Informasi Dataset yang Digunakan

a. Airline Passenger Satisfaction

Customer Type	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	On-board service	Leg room	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction
Loyal Customer	Personal Travel	Eco Plus	460	3	4	3	1	5	3	5	5	4	3	4	4	5	5	25	18	neutral or dissatisfied
disloyal Customer	Business travel	Business	235	3	2	3	3	1	3	1	1	1	5	3	1	4	1	1	6	neutral or dissatisfied
Loyal Customer	Business travel	Business	1142	2	2	2	2	5	5	5	5	4	3	4	4	4	5	0	0	satisfied
Loyal Customer	Business travel	Business	562	2	5	5	5	2	2	2	2	2	5	3	1	4	2	11	9	neutral or dissatisfied
Loyal Customer	Business travel	Business	214	3	3	3	3	4	5	5	3	3	4	4	3	3	3	0	0	satisfied
Loyal Customer	Personal Travel	Eco	1180	3	4	2	1	1	2	1	1	3	4	4	4	4	1	0	0	neutral or dissatisfied
Loyal Customer	Personal Travel	Eco	1276	2	4	2	3	2	2	2	2	3	3	4	3	5	2	9	23	neutral or dissatisfied
Loyal Customer	Business travel	Business	2035	4	3	4	4	5	5	5	5	5	5	5	4	5	4	4	0	satisfied
Loyal Customer	Business travel	Business	853	1	2	2	2	4	3	3	1	1	2	1	4	1	2	0	0	neutral or dissatisfied
disloyal Customer	Business travel	Eco	1061	3	3	3	4	2	3	3	2	2	3	4	4	3	2	0	0	neutral or dissatisfied

b. IRIS

sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
7	3.2	4.7	1.4	Iris-versicolor
6.4	3.2	4.5	1.5	Iris-versicolor
6.9	3.1	4.9	1.5	Iris-versicolor
6.7	3.3	5.7	2.1	Iris-virginica
7.2	3.2	6	1.8	Iris-virginica
6.2	2.8	4.8	1.8	Iris-virginica

c. TIC TAC TU

Top-Left-Square	Top-Middle-Square	Top-Right-Square	Middle-Left-Square	Middle-Middle-Square	Middle-Right-Square	Bottom-Left-Square	Bottom-Middle-Square	Bottom-Right-Square	Class
b'x'	b'x'	b'x'	b'x'	b'o'	b'o'	b'x'	b'o'	b'o'	b'positive'
b'x'	b'x'	b'x'	b'x'	b'o'	b'o'	b'o'	b'x'	b'o'	b'positive'
b'x'	b'x'	b'x'	b'x'	b'o'	b'o'	b'o'	b'o'	b'x'	b'positive'
b'x'	b'x'	b'x'	b'x'	b'o'	b'o'	b'o'	b'b'	b'b'	b'positive'
b'x'	b'x'	b'x'	b'x'	b'o'	b'o'	b'b'	b'o'	b'b'	b'positive'

Top-Left-Square	Top-Middle-Square	Top-Right-Square	Middle-Left-Square	Middle-Middle-Square	Middle-Right-Square	Bottom-Left-Square	Bottom-Middle-Square	Bottom-Right-Square	Class
b'x'	b'x'	b'o'	b'x'	b'x'	b'o'	b'o'	b'b'	b'o'	b'negative'
b'x'	b'x'	b'o'	b'x'	b'x'	b'o'	b'b'	b'o'	b'o'	b'negative'
b'x'	b'x'	b'o'	b'x'	b'x'	b'b'	b'o'	b'o'	b'o'	b'negative'
b'x'	b'x'	b'o'	b'x'	b'o'	b'x'	b'o'	b'o'	b'b'	b'negative'
b'x'	b'x'	b'o'	b'x'	b'o'	b'x'	b'o'	b'b'	b'o'	b'negative'

d. Water_Potability

ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
8.316766	214.3734	22018.42	8.059332	356.8861	363.2665	18.43652	100.3417	4.628771	Aman
9.092223	181.1015	17978.99	6.5466	310.1357	398.4108	11.55828	31.99799	4.075075	Aman
5.584087	188.3133	28748.69	7.544869	326.6784	280.4679	8.399735	54.91786	2.559708	Aman
10.22386	248.0717	28749.72	7.513408	393.6634	283.6516	13.7897	84.60356	2.672989	Aman
8.635849	203.3615	13672.09	4.563009	303.3098	474.6076	12.36382	62.79831	4.401425	Aman
11.18028	227.2315	25484.51	9.0772	404.0416	563.8855	17.92781	71.9766	4.370562	Aman

Lampiran 2. Hasil Perhitungan Manual

Tabel 1. Tabel Data *Dummy*

sepal_length	sepal_width	petal_length	petal_width	species
5.2	4.1	1.5	0.2	Iris-versicolor
5.2	4.1	1.5	0.2	Iris-setosa
7	3.2	4.5	1.5	Iris-versicolor
6.5	3.2	4.5	1.5	Iris-versicolor
4.7	3.2	1.6	0.2	Iris-setosa
5.5	2.3	4.5	1.5	Iris-versicolor
6.5	2.8	1.5	1.5	Iris-versicolor
5.2	3.4	1.5	0.4	Iris-setosa
7	3.1	4.9	1.5	Iris-setosa
4.7	3.1	1.6	0.2	Iris-versicolor

Data pada Tabel 1 adalah data *dummy*. Algoritma dari *C* 4.5 dan *C* 4.5 berbasis *Adaboost* akan diterapkan pada data tersebut. Tahap pertama adalah inisiasi bobot dengan rumus $(x) = \frac{1}{N} = \frac{1}{10} = 0,1$. hasil inisiasi bobot dapat dilihat seperti pada Tabel 2.

Tabel 2. Tabel Data dengan bobot

sepal_length	sepal_width	petal_length	petal_width	species	bobot
5.2	4.1	1.5	0.2	Iris-versicolor	0.1
5.2	4.1	1.5	0.2	Iris-setosa	0.1
7	3.2	4.5	1.5	Iris-versicolor	0.1
6.5	3.2	4.5	1.5	Iris-versicolor	0.1
4.7	3.2	1.6	0.2	Iris-setosa	0.1
5.5	2.3	4.5	1.5	Iris-versicolor	0.1
6.5	2.8	1.5	1.5	Iris-versicolor	0.1
5.2	3.4	1.5	0.4	Iris-setosa	0.1
7	3.1	4.9	1.5	Iris-setosa	0.1
4.7	3.1	1.6	0.2	Iris-versicolor	0.1

Iterasi pertama adalah proses algoritma *C* 4.5. Bobot dapat diabaikan pada perhitungan ini. Tahap pertama adalah mencari gain dari masing-masing atribut.

Tertera pada Tabel 3.

Tabel 3. Tabel Menghitung Gain Node 1

Node		Jumlah Kasus	Iris-	Iris-	Entropy	Total	Gain
			setosa	versicolor			
	TOTAL	10	4	6	0.97095059		
	G1	sepal_length					0.295462
		4.7	2	1	1	0.2	
		5.2	3	2	1	0.91829583	0.275489
		5.5	1	0	1	0	0
		6.5	2	0	2	0	0
		7	2	1	1	1	0.2
	G2	sepal_width					0.295462
		2.3	1	0	1	0	0
		2.8	1	0	1	0	0
		3.1	2	1	1	1	0.2
		3.2	3	1	2	0.91829583	0.275489
		3.4	1	1	0	0	0
		4.1	2	1	1	1	0.2
	G3	petal_length					0.370951
		1.5	4	2	2	1	0.4
		1.6	2	1	1	1	0.2
		4.5	3	0	3	0	0
		4.9	1	1	0	0	0
	G4	petal_width					0.209987
		0.2	4	2	2	1	0.4
		0.4	1	1	0	0	0
		1.5	5	1	4	0.72192809	0.360964

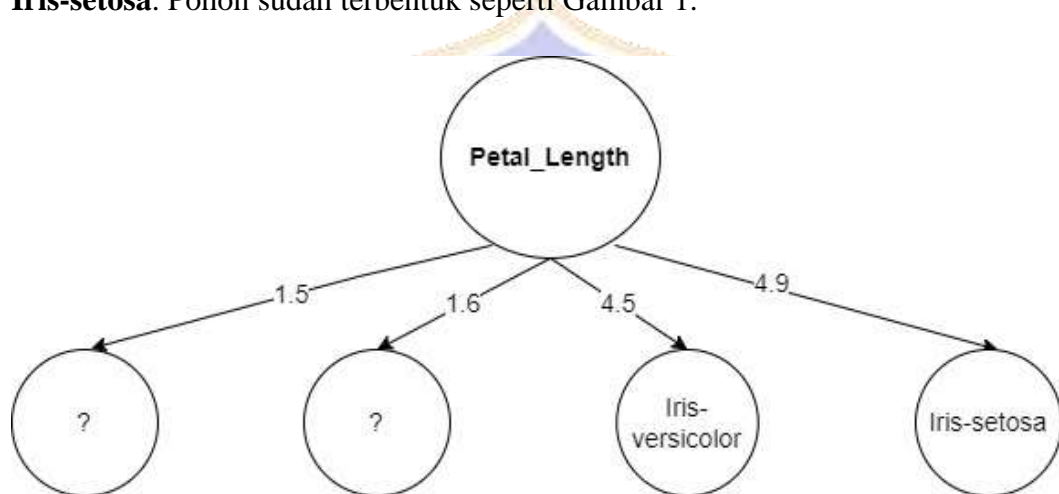
Entropy dihitung menggunakan rumus $Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$

Kolom **TOTAL** dihitung menggunakan potongan rumus gain, yaitu $Total = \frac{|S_i|}{|S|} * Entropy(S_i)$

$Entropy(S_i)$

Gain dihitung menggunakan rumus $Gain(S, A) = Entropy(S) - \sum_{i=1}^n Total(i)$

Gain tertinggi didapatkan pada atribut **petal_length** yaitu 0.370951. Atribut ini menjadi node pertama pada pohon yang akan dihasilkan. Data yang bernilai “4.5” merujuk pada hasil **Iris-versicolor** dan yang memiliki nilai “4.9” merujuk nilai **Iris-setosa**. Pohon sudah terbentuk seperti Gambar 1.



Gambar 1. Gambar Iterasi 1 Node 1

Proses selanjutnya menghitung data yang bernilai “1.5” dan “1.6”. Data yang bernilai “1.5” dapat dilihat pada Tabel 4.

Tabel 4. Tabel Iterasi 1 Petal_Length 1.5

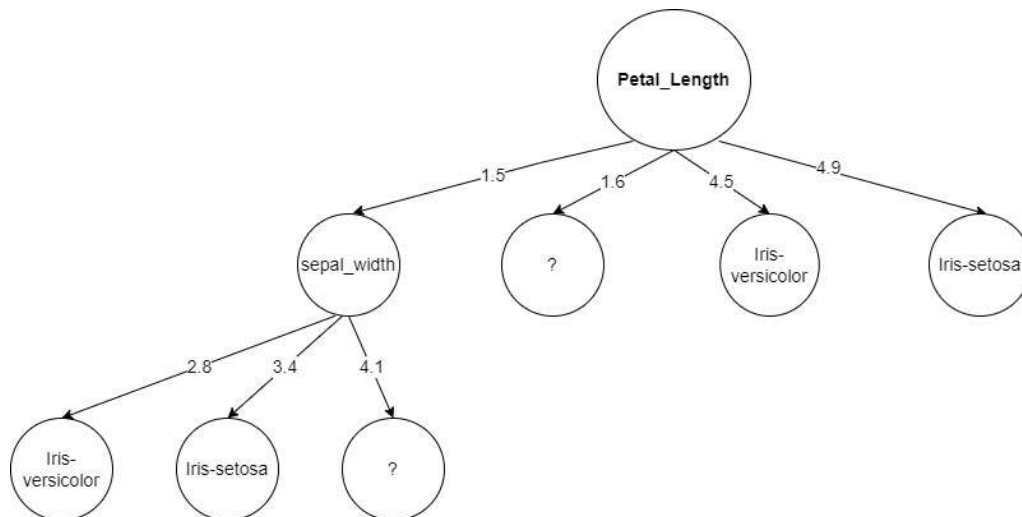
sepal_length	sepal_width	petal_length	petal_width	species
5.2	4.1	1.5	0.2	Iris-versicolor
5.2	4.1	1.5	0.2	Iris-setosa
6.5	2.8	1.5	1.5	Iris-versicolor
5.2	3.4	1.5	0.4	Iris-setosa

Selanjutnya menghitung gain dari tabel 4. Perhitungan gain dapat dilihat seperti pada Tabel 5.

Tabel 5. Tabel Menghitung Gain Node 2

Node			Jumlah Kasus	Iris-setosa	Iris-versicolor	Entropy	Total	Gain
	TOTAL		4	2	2	1		
	G1	sepal_length						0.311278
		4.7	0	0	0	0	0	
		5.2	3	2	1	0.91829583	0.688722	
		5.5	0	0	0	0	0	
		6.5	1	0	1	0	0	
		7	0	0	0	0	0	
	G2	sepal_width						0.5
		2.3	0	0	0	0	0	
		2.8	1	0	1	0	0	
		3.1	0	0	0	0	0	
		3.2	0	0	0	0	0	
		3.4	1	1	0	0	0	
		4.1	2	1	1	1	0.5	
	G4	petal_width						0.5
		0.2	2	1	1	1	0.5	
		0.4	1	1	0	0	0	
		1.5	1	0	1	0	0	

Berdasarkan tabel 3.16, gain tertinggi terdapat pada atribut sepal_width. Data yang bernilai “2.8” merujuk pada Iris-versicolor dan data bernilai “3.4” merujuk pada Iris-setosa. Node baru ditambahkan pada pohon. Pohon tersebut dapat dilihat pada Gambar 2.



Gambar 2. Gambar Iterasi 1 Node 2

Selanjutnya menghitung data yang bernilai “4.1”. Data dapat dilihat pada tabel 3.17.

Tabel 6. Tabel Data Sepal_width = “4.1”

sepal_length	sepal_width	petal_length	petal_width	species
5.2	4.1	1.5	0.2	Iris-versicolor
5.2	4.1	1.5	0.2	Iris-setosa

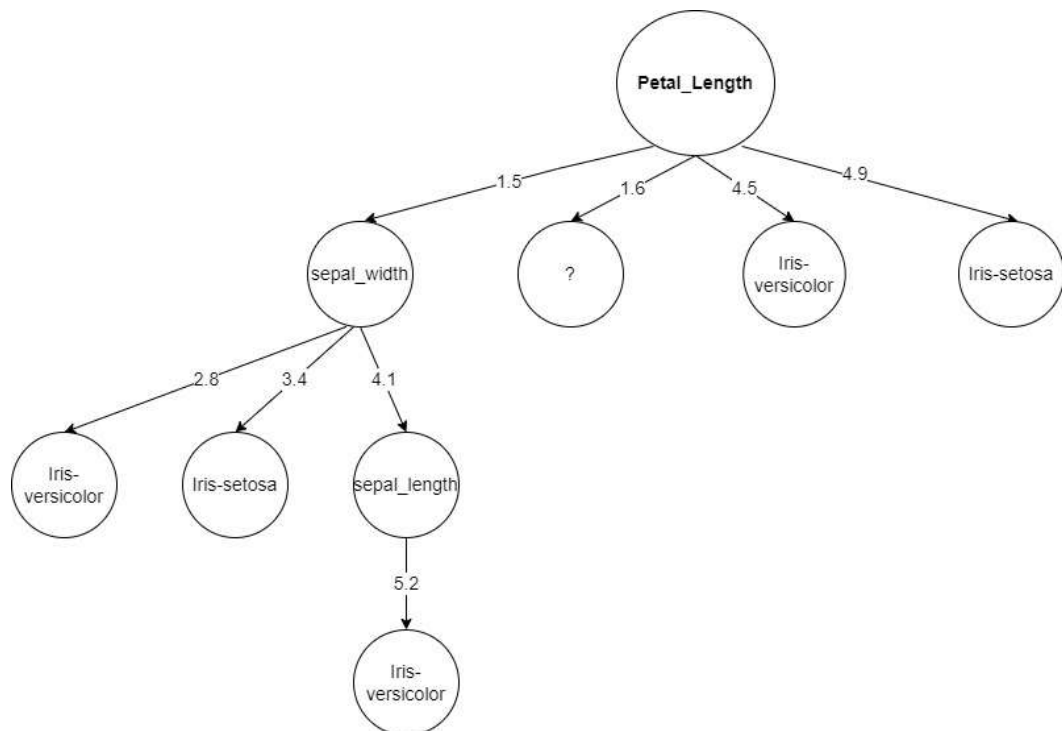
Dilanjutkan dengan menghitung gain dari tabel 7. Perhitungan dapat dilihat seperti pada Tabel 7.

Tabel 8. Tabel Menghitung Gain Node 3

Node			Jumlah Kasus	Iris-		Entropy	Total	Gain
				setosa	versicolor			
	TOTAL		2	1	1	1		
	G1	sepal_length						1
		4.7	0	0	0	0	0	
		5.2	1	1	1	0	0	
		5.5	0	0	0	0	0	
		6.5	0	0	0	0	0	

		7	0	0	0	0	0	
--	--	---	---	---	---	---	---	--

Tabel 8 menunjukkan bahwa, data “5.2” memiliki nilai yang sama yaitu “1”. Pada kasus seperti ini, pilih secara acak baik Iris-setosa atau Iris-versicolor. Misalnya dipilih jawaban “Iris-versicolor”. Untuk node **Petal_length** dengan nilai “1.5” sudah berakhir. Gambar pohon menjadi seperti Gambar 3.



Gambar 3. Gambar Iterasi 1 Node 3

Dilanjutkan dengan menghitung node Petal_length dengan nilai “1.6”. Data dapat dilihat pada Tabel 9.

Tabel 9. Tabel Data Petal_length = “1.6”

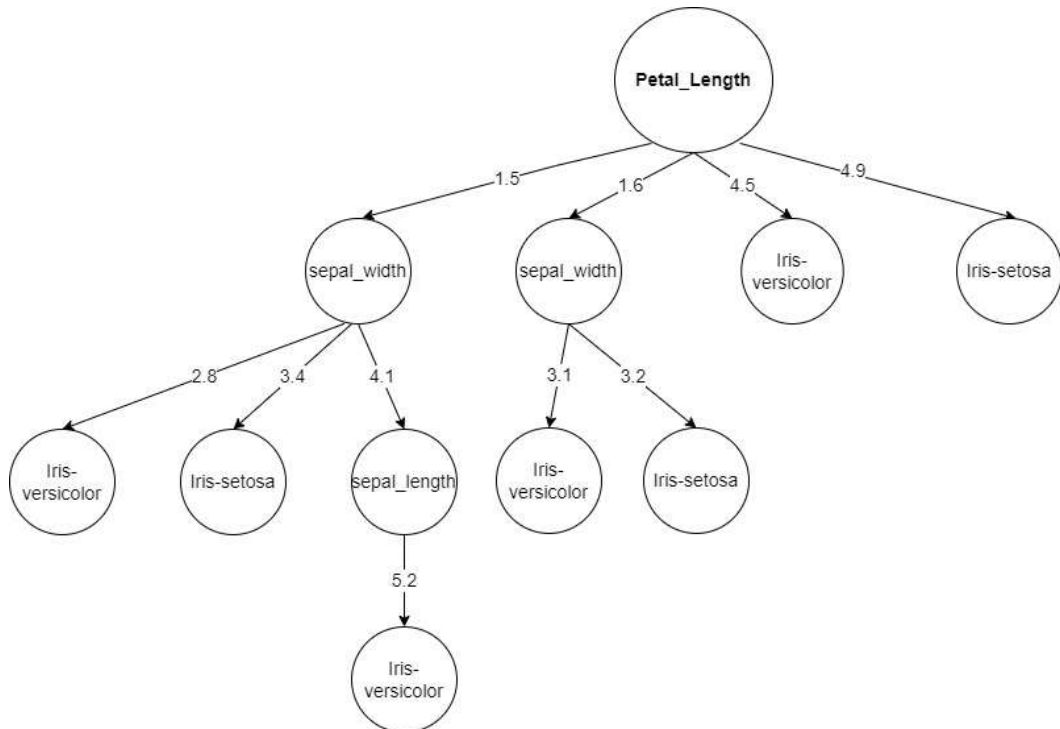
sepal_length	sepal_width	petal_length	petal_width	species
4.7	3.2	1.6	0.2	Iris-setosa
4.7	3.1	1.6	0.2	Iris-versicolor

Dilanjutkan dengan mencari gain dari tabel 3.19. Perhitungan dapat dilihat seperti pada tabel 10.

Tabel 10. Tabel Menghitung Gain Node 4

Node			Jumlah Kasus	Iris-setosa	Iris-versicolor	Entropy	Total	Gain
	TOTAL		2	1	1	1		
	G1	sepal_length						0
		4.7	2	1	1	1	1	
		5.2	0	0	0	0	0	
		5.5	0	0	0	0	0	
		6.5	0	0	0	0	0	
		7	0	0	0	0	0	
	G2	sepal_width						1
		2.3	0	0	0	0	0	
		2.8	0	0	0	0	0	
		3.1	1	0	1	0	0	
		3.2	1	1	0	0	0	
		3.4	0	0	0	0	0	
		4.1	0	0	0	0	0	
	G4	petal_width						0
		0.2	2	1	1	1	1	
		0.4	0	0	0	0	0	
		1.5	0	0	0	0	0	

Berdasarkan Tabel 10, masing-masing gain memiliki nilai 0 dan sepal_width memiliki nilai gain 1. Hal ini menjadikan sepal_width dipilih menjadi atribut pada node selanjutnya. Gambar pohon menjadi seperti Gambar 4.



Gambar 4. Gambar Iterasi 1 Node 4

Pohon diubah menjadi bentuk hukum implikasi. Hukum ini selanjutnya di ujikan ke data awal seperti pada Tabel 11

Tabel 11. Tabel Hasil Pengujian Pohon Pada Data Awal

sepal_length	sepal_width	petal_length	petal_width	species	prediksi pohon	bobot
5.2	4.1	1.5	0.2	Iris-versicolor	Iris-versicolor	0.1
5.2	4.1	1.5	0.2	Iris-setosa	Iris-versicolor	0.1
7	3.2	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1
6.5	3.2	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1
4.7	3.2	1.6	0.2	Iris-setosa	Iris-setosa	0.1
5.5	2.3	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1
6.5	2.8	1.5	1.5	Iris-versicolor	Iris-versicolor	0.1
5.2	3.4	1.5	0.4	Iris-setosa	Iris-setosa	0.1
7	3.1	4.9	1.5	Iris-setosa	Iris-setosa	0.1
4.7	3.1	1.6	0.2	Iris-versicolor	Iris-versicolor	0.1

- Algoritma Berbasis Adaboost

Hasil prediksi pohon memiliki 1 buah prediksi yang bernilai salah. Kesalahan prediksi pohon digunakan untuk menghitung *error rate* dari pohon. *Error rate* dihitung menggunakan rumus

$$\varepsilon_t = \sum_{i=1}^N D(x_i) \cdot I [y_i \neq h_t(x_i)]$$

Rumus ini menjabarkan bahwa error rate adalah jumlah bobot dari data yang diprediksi salah. Error rate dari pohon ini adalah

$$\varepsilon_t = \frac{1}{10} = 0,1$$

Proses selanjutnya adalah menghitung bobot pohon. Bobot pohon dihitung menggunakan rumus

$$\alpha_t = \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) = \log\left(\frac{1-0.3}{0.3}\right) = 0.954$$

Proses selanjutnya adalah *boosting* bobot dari data yang diprediksi salah. Proses *boosting* menggunakan rumus

$$D(x) = D(x) \cdot \exp(\alpha)$$

Misalnya pada data nomor 2, maka proses *boosting* akan menjadi seperti berikut

$$D(x) = 0,1 \cdot \exp 0,954 = 0,25967$$

Seluruh bobot data yang diprediksi salah, diterapkan proses *boosting*. Tabel 3.22 menjabarkan hasil dari *boosting* pada masing-masing bobot.

Tabel 12. Tabel Hasil *boosting* bobot iterasi 1

Sepal _length	Sepal _width	Petal _length	Petal _width	species	prediksi pohon	bobot awal	Bobot baru
5.2	4.1	1.5	0.2	Iris-versicolor	Iris-versicolor	0.1	0.1
5.2	4.1	1.5	0.2	Iris-setosa	Iris-versicolor	0.1	0.259670286
7	3.2	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1

Sepal _length	Sepal _width	Petal _length	Petal _width	species	prediksi pohon	bobot awal	Bobot baru
6.5	3.2	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1
4.7	3.2	1.6	0.2	Iris-setosa	Iris-setosa	0.1	0.1
5.5	2.3	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1
6.5	2.8	1.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1
5.2	3.4	1.5	0.4	Iris-setosa	Iris-setosa	0.1	0.1
7	3.1	4.9	1.5	Iris-setosa	Iris-setosa	0.1	0.1
4.7	3.1	1.6	0.2	Iris-versicolor	Iris-versicolor	0.1	0.1
					total	1	1.159670286

Jumlah seluruh bobot harus bernilai 1. Jumlah bobot baru pada tabel di atas masih belum bernilai 1, sehingga perlu dilakukan normalisasi pada seluruh bobot. Proses normalisasi menggunakan rumus $D(x) = \frac{D(x)}{D_{total}}$. Hasil normalisasi dapat dilihat pada Tabel 13.

Tabel 13. Tabel Hasil normalisasi bobot iterasi 1

sepal_ length	sepal_ width	petal_ length	petal_ width	species	prediksi pohon	bobot awal	Bobot baru	Bobot Final
5.2	4.1	1.5	0.2	Iris-versicolor	Iris-versicolor	0.1	0.1	0.08623
5.2	4.1	1.5	0.2	Iris-setosa	Iris-versicolor	0.1	0.2597	0.22392
7	3.2	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1	0.08623
6.5	3.2	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1	0.08623
4.7	3.2	1.6	0.2	Iris-setosa	Iris-setosa	0.1	0.1	0.08623
5.5	2.3	4.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1	0.08623
6.5	2.8	1.5	1.5	Iris-versicolor	Iris-versicolor	0.1	0.1	0.08623
5.2	3.4	1.5	0.4	Iris-setosa	Iris-setosa	0.1	0.1	0.08623
7	3.1	4.9	1.5	Iris-setosa	Iris-setosa	0.1	0.1	0.08623
4.7	3.1	1.6	0.2	Iris-versicolor	Iris-versicolor	0.1	0.1	0.08623
					total	1	1.1597	1

Iterasi dilanjutkan. Data dari tabel 13 digunakan pada iterasi selanjutnya. Langkah-langkah dari perhitungan di iterasi 2 dan seterusnya hampir sama dengan iterasi 1. Namun perhitungan pada iterasi 2 dan seterusnya sudah menggunakan bobot, baik dalam mencari Entropy maupun mencari Gain. Data pertama pada iterasi 2 dapat dilihat pada Tabel 14.

Tabel 14. Tabel Data awal iterasi 2

Sepal _length	Sepal _width	Petal _length	Petal _width	species	Bobot Final
5.2	4.1	1.5	0.2	Iris-versicolor	0.086231
5.2	4.1	1.5	0.2	Iris-setosa	0.223917
7	3.2	4.5	1.5	Iris-versicolor	0.086231
6.5	3.2	4.5	1.5	Iris-versicolor	0.086231
4.7	3.2	1.6	0.2	Iris-setosa	0.086231
5.5	2.3	4.5	1.5	Iris-versicolor	0.086231
6.5	2.8	1.5	1.5	Iris-versicolor	0.086231
5.2	3.4	1.5	0.4	Iris-setosa	0.086231
7	3.1	4.9	1.5	Iris-setosa	0.086231
4.7	3.1	1.6	0.2	Iris-versicolor	0.086231

Tahap pertama dari iterasi 2 adalah mencari gain tertinggi dari beberapa atribut. Atribut yang seluruh nilainya sama dapat diabaikan. Perhitungan dapat dilihat pada tabel 15.

Tabel 15. Tabel Menghitung Gain Iterasi 2 Node 1

Node		Jumlah Kasus	Iris- setosa	Iris- versicolor	Entropy	Total	Gain
	TOTAL	10	0.482612	0.517388	0.43210555		
	G3	sepal_length					0.129963
		4.7	2	0.086231	0.086231	0.39111486	0.078223
		5.2	3	0.310149	0.086231	0.48565413	0.145696

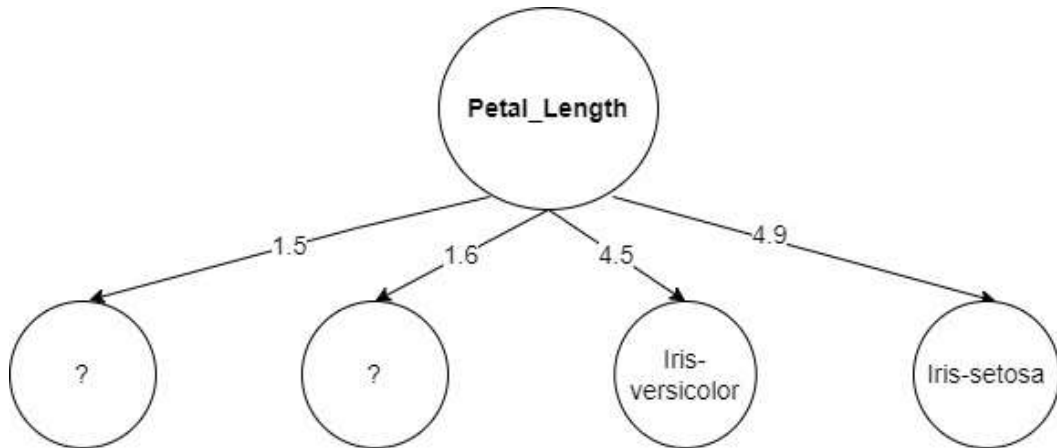
		5.5	1	0	0.086231	0	0	
		6.5	2	0	0.172463	0	0	
		7	2	0.086231	0.086231	0.39111486	0.078223	
	G2	sepal_width						0.128816
		2.3	1	0	0.086231	0	0	
		2.8	1	0	0.086231	0	0	
		3.1	2	0.086231	0.086231	0.39111486	0.078223	
		3.2	3	0.086231	0.172463	0.38406939	0.115221	
		3.4	1	0.086231	0	0	0	
		4.1	2	0.223917	0.086231	0.54923059	0.109846	
	G3	petal_length						0.161247
		1.5	4	0.310149	0.172463	0.48158961	0.192636	
		1.6	2	0.086231	0.086231	0.39111486	0.078223	
		4.5	3	0	0.258694	0	0	
		4.9	1	0.086231	0	0	0	
	G4	petal_width						0.055902
		0.2	4	0.310149	0.172463	0.48158961	0.192636	
		0.4	1	0.086231	0	0	0	
		1.5	5	0.086231	0.344926	0.36713633	0.183568	

Perbedaan perhitungan menggunakan bobot terletak di *entropy*. *Entropy* dihitung menggunakan rumus

$$Entropy(S) = \sum_{i=1}^n -\rho_i * \log_2 \rho_i$$

Nilai ρ_i yang awalnya adalah proporsi S_i terhadap S , diubah menjadi total bobot S_i .

Misalnya pada data `petal_length = "1.5"`. Total bobot yang bernilai Iris-setosa adalah 0,310149. Total bobot yang bernilai Iris-versicolor adalah $\rho_i = 0.0862 + 0.0862$ (terdapat pada data nomor 1 dan 7). Perhitungan *entropy* menggunakan nilai tersebut. Diperoleh nilai gain tertinggi pada atribut `petal_length`. Gambar pohon hasil dari perhitungan tabel 14 dapat dilihat pada Gambar 5.



Gambar 5. Gambar Iterasi 2 node 1

Proses selanjutnya menghitung data yang bernilai “1.5” dan “1.6”. Dihitung data dengan nilai “1.5” terlebih dahulu. Data tersebut dapat dilihat pada Tabel 16.

Tabel 16. Tabel Data petal_length = “1.5”

Sepal _length	Sepal _width	Petal _length	Petal _width	species	Bobot Final
5.2	4.1	1.5	0.2	Iris-versicolor	0.086231
5.2	4.1	1.5	0.2	Iris-setosa	0.223917
6.5	2.8	1.5	1.5	Iris-versicolor	0.086231
5.2	3.4	1.5	0.4	Iris-setosa	0.086231

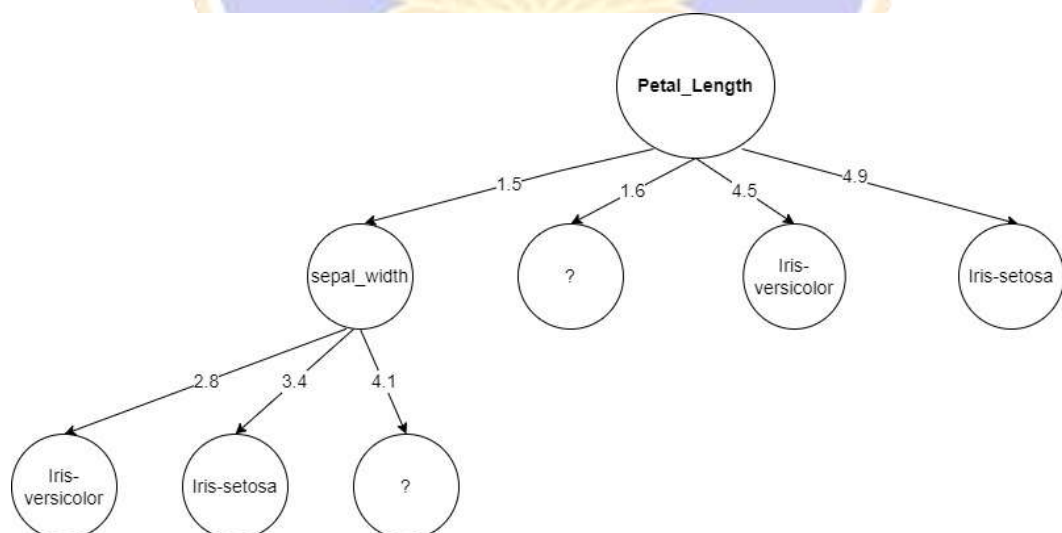
Selanjutnya menghitung gain dari tabel 16. Perhitungan gain dapat dilihat seperti pada tabel 17.

Tabel 17. Tabel Menghitung Gain Iterasi 2 Node 2

Node			Jumlah Kasus	Iris- setosa	Iris- versicolor	Entropy	Total	Gain
	TOTAL		4	0.310149	0.172463	0.48158961		
	G3	sepal_length						0.117349
		4.7	0	0	0	0	0	
		5.2	3	0.310149	0.086231	0.48565413	0.364241	

		5.5	0	0	0	0	0	
		6.5	1	0	0.086231	0	0	
		7	0	0	0	0	0	
	G2	sepal_width						0.206974
		2.3	0	0	0	0	0	
		2.8	1	0	0.086231	0	0	
		3.1	0	0	0	0	0	
		3.2	0	0	0	0	0	
		3.4	1	0.086231	0	0	0	
		4.1	2	0.223917	0.086231	0.54923059	0.274615	
	G4	petal_width						0.206974
		0.2	2	0.223917	0.086231	0.54923059	0.274615	
		0.4	1	0.086231	0	0	0	
		1.5	1	0	0.086231	0	0	

Gain tertinggi terdapat pada atribut sepal_width. Data yang bernilai “2.8” merujuk pada Iris-versicolor, data yang bernilai “3.4” merujuk pada iris-setosa dan data bernilai “4.1” merujuk pada keduanya. Node baru ditambahkan pada pohon, sehingga pohon seperti Gambar 6.



Gambar 6. Gambar Iterasi 2 node 2

Selanjutnya menghitung data yang bernilai “4.1”. Data dapat dilihat seperti pada Tabel 3.28.

Tabel 18. Tabel Data Sepal_width = “4.1”

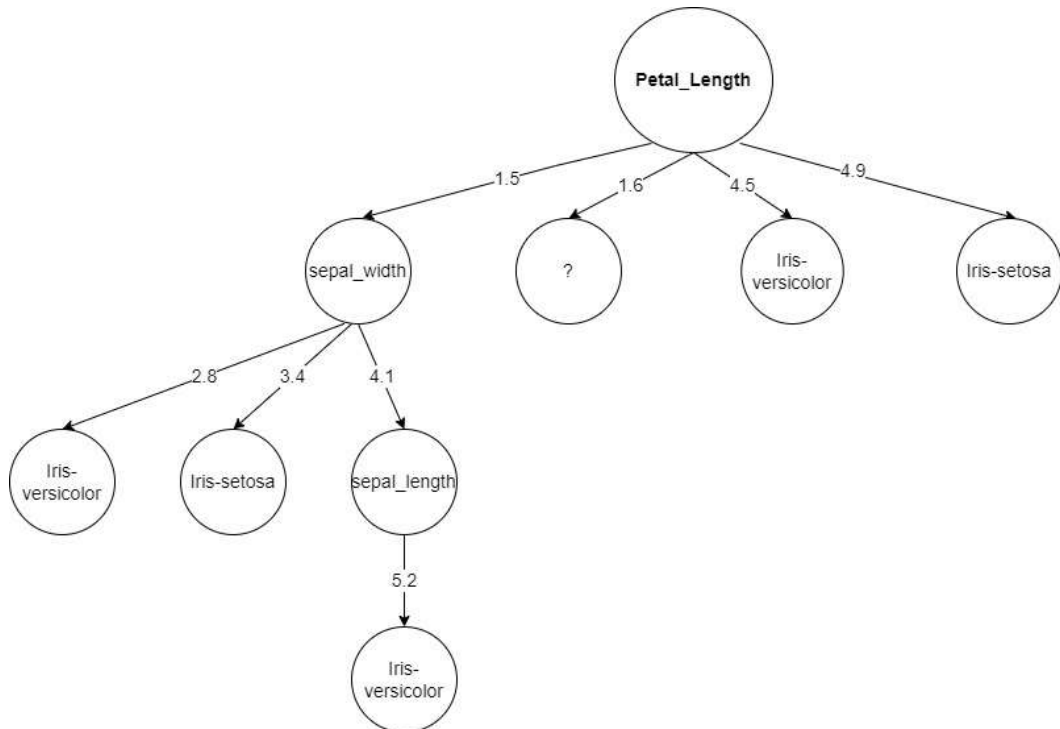
sepal_length	sepal_width	petal_length	petal_width	species	Bobot Final
5.2	4.1	1.5	0.2	Iris-versicolor	0.086231
5.2	4.1	1.5	0.2	Iris-setosa	0.223917

Dilanjutkan dengan menghitung gain dari tabel 3.28. Perhitungan dapat seperti dilihat pada Tabel 19.

Tabel 19. Tabel Menghitung Gain Iterasi 2 Node 3

Node			Jumlah Kasus	Iris-setosa	Iris-versicolor	Entropy	Total	Gain
	TOTAL		2	0.223917	0.086231	0.54923059		
	G3	sepal_length						0
		4.7	0	0	0	0	0	
		5.2	2	0.223917	0.086231	0.54923059	0.549231	
		5.5	0	0	0	0	0	
		6.5	0	0	0	0	0	
		7	0	0	0	0	0	

Berdasarkan Tabel 19, sepal_length adalah atribut terakhir pada perhitungan ini. Pemilihan jawaban pada kasus seperti ini dapat menggunakan bobot tertinggi. Bobot tertinggi untuk Data bernilai “5.2” adalah “Iris-versicolor” dengan bobot 0.086231, lebih besar dari jawaban “Iris-setosa”. Untuk node Sepal_length dengan nilai “5.2” sudah berakhir. Gambar pohon ditunjukkan pada Gambar 7.



Gambar 7. Gambar Iterasi 2 node 3

Dilanjutkan dengan menghitung node Petal_length dengan nilai “1.6”. Data dapat dilihat pada Tabel 3.30.

Tabel 20. Tabel Data Petal_length = “1.6”

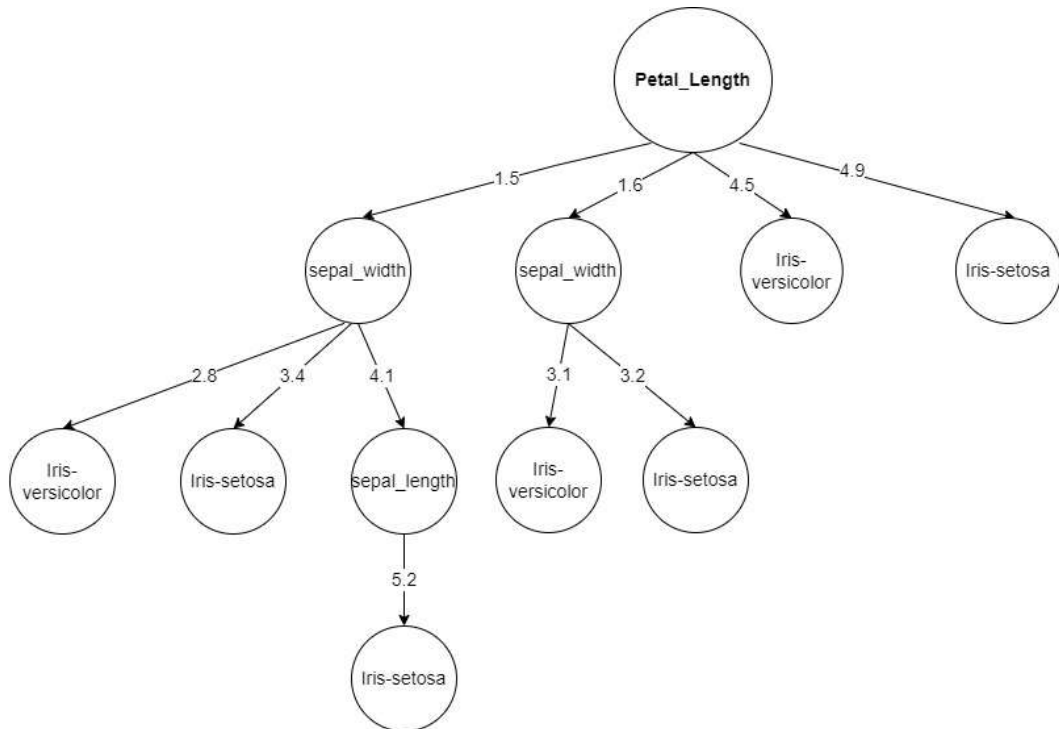
sepal_length	sepal_width	petal_length	petal_width	species	Bobot Final
4.7	3.2	1.6	0.2	Iris-setosa	0.086231
4.7	3.1	1.6	0.2	Iris-versicolor	0.086231

Dilanjutkan dengan mencari gain dari tabel 20. Perhitungan dapat dilihat seperti pada Tabel 21.

Tabel 21. Tabel Menghitung Gain Iterasi 2 Node 4

Node			Jumlah Kasus	Iris-setosa	Iris-versicolor	Entropy	Total	Gain
	TOTAL		2	0.086231	0.086231	0.39111486		
	G3	sepal_length						0
		4.7	2	0.086231	0.086231	0.39111486	0.391115	
		5.2	0	0	0	0	0	
		5.5	0	0	0	0	0	
		6.5	0	0	0	0	0	
		7	0	0	0	0	0	
	G2	sepal_width						0.391115
		2.3	0	0	0	0	0	
		2.8	0	0	0	0	0	
		3.1	1	0	0.086231	0	0	
		3.2	1	0.086231	0	0	0	
		3.4	0	0	0	0	0	
		4.1	0	0	0	0	0	
	G4	petal_width						0
		0.2	2	0.086231	0.086231	0.39111486	0.391115	
		0.4	0	0	0	0	0	
		1.5	0	0	0	0	0	

Berdasarkan perhitungan pada Tabel 21, dipilih atribut sepal_width sebagai node selanjutnya. Gambar pohon ditunjukkan oleh Gambar 8.



Gambar 8. Gambar Iterasi 2 node 4

Pohon diubah menjadi bentuk hukum implikasi. Hukum ini selanjutnya di ujikan ke data awal seperti pada Tabel 22.

Tabel 22. Tabel Hasil pengujian pohon iterasi 2

Sepal _length	Sepal _width	Petal _length	Petal _width	species	Bobot Final	Prediksi Pohon
5.2	4.1	1.5	0.2	Iris-versicolor	0.086231	Iris-setosa
5.2	4.1	1.5	0.2	Iris-setosa	0.223917	Iris-setosa
7	3.2	4.5	1.5	Iris-versicolor	0.086231	Iris-versicolor
6.5	3.2	4.5	1.5	Iris-versicolor	0.086231	Iris-versicolor
4.7	3.2	1.6	0.2	Iris-setosa	0.086231	Iris-setosa
5.5	2.3	4.5	1.5	Iris-versicolor	0.086231	Iris-versicolor
6.5	2.8	1.5	1.5	Iris-versicolor	0.086231	Iris-versicolor
5.2	3.4	1.5	0.4	Iris-setosa	0.086231	Iris-setosa
7	3.1	4.9	1.5	Iris-setosa	0.086231	Iris-setosa
4.7	3.1	1.6	0.2	Iris-versicolor	0.086231	Iris-versicolor

Hasil prediksi pohon tidak memiliki prediksi yang bernilai salah. Kesalahan prediksi pohon digunakan untuk menghitung *error rate* dari pohon.

$$\varepsilon_t = 0$$

Proses selanjutnya adalah menghitung bobot pohon. Bobot pohon dihitung menggunakan rumus

$$\alpha_t = \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) = \log\left(\frac{1-0}{0}\right) = 1$$

Proses selanjutnya adalah *boosting* bobot dari data yang diprediksi salah. Proses *boosting* menggunakan rumus

$$D(x) = D(x) \cdot \exp(\alpha)$$

Misalnya pada data nomor 1, maka proses *boosting* akan menjadi seperti berikut

$$D(x) = 0,0882 \cdot \exp 1 = 0,234401$$

Seluruh bobot data yang diprediksi salah, diterapkan proses *boosting*. Tabel 3.33 menunjukkan hasil dari *boosting* pada masing-masing bobot. Proses selanjutnya adalah normalisasi bobot.

Tabel 23. Tabel Hasil *boosting* bobot iterasi 2

Sepal _length	Sepal _width	Petal _length	Petal _width	Species	prediksi pohon	Bobot awal	Bobot baru	Bobot Final	Boosting	Bobot Normalisasi
5.2	4.1	1.5	0.2	Iris- versicolor	Iris- setosa	0.1	0.1	0.086231	0.310149	0.282440329
5.2	4.1	1.5	0.2	Iris- setosa	Iris- setosa	0.1	0.25967	0.223917	0.193087	0.175836699
7	3.2	4.5	1.5	Iris- versicolor	Iris- versicolor	0.1	0.1	0.086231	0.074359	0.067715371
6.5	3.2	4.5	1.5	Iris- versicolor	Iris- versicolor	0.1	0.1	0.086231	0.074359	0.067715371
4.7	3.2	1.6	0.2	Iris- setosa	Iris- setosa	0.1	0.1	0.086231	0.074359	0.067715371
5.5	2.3	4.5	1.5	Iris- versicolor	Iris- versicolor	0.1	0.1	0.086231	0.074359	0.067715371
6.5	2.8	1.5	1.5	Iris- versicolor	Iris- versicolor	0.1	0.1	0.086231	0.074359	0.067715371
5.2	3.4	1.5	0.4	Iris- setosa	Iris- setosa	0.1	0.1	0.086231	0.074359	0.067715371

7	3.1	4.9	1.5	Iris-setosa	Iris-setosa	0.1	0.1	0.086231	0.074359	0.067715371
4.7	3.1	1.6	0.2	Iris-versicolor	Iris-versicolor	0.1	0.1	0.086231	0.074359	0.067715371
						total	1	1.15967	1.098105	1

Iterasi berlanjut sesuai banyak iterasi yang ditentukan. Proses selanjutnya adalah mengukur akurasi dengan *confusion matrix*. Hukum yang dihasilkan dari kedua algoritma, diuji pada data uji yang telah disiapkan. Tabel 24 menunjukkan data uji.

- Evaluasi

Tabel 24. Tabel Data Uji

No	sepal_length	sepal_width	petal_length	petal_width	species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-versicolor
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.2	4.1	1.5	0.2	Iris-versicolor
7	5.2	4.1	1.5	0.2	Iris-setosa
8	7	3.2	4.5	1.5	Iris-versicolor
9	6.5	3.2	4.5	1.5	Iris-versicolor
10	4.7	3.2	1.6	0.2	Iris-setosa
11	5.5	2.3	4.5	1.5	Iris-versicolor
12	6.5	2.8	1.5	1.5	Iris-versicolor
13	5.2	3.4	1.5	0.4	Iris-setosa
14	7	3.1	4.9	1.5	Iris-setosa
15	4.7	3.1	1.6	0.2	Iris-versicolor

Data uji ini digunakan untuk menguji akurasi dari masing-masing pohon. Setiap hukum yang tercipta dari masing-masing pohon akan di tuangkan ke data uji. Untuk pohon C4.5 akan langsung menghasilkan jawaban “Iris-setosa” atau “Iris-versicolor”. Pohon C4.5 berbasis Adaboost akan menghasilkan beberapa jawaban sehingga perlu proses pengolahan kembali. Setiap pohon yang dihasilkan oleh iterasi Adaboost akan memberikan bobot pada pohon tersebut. Bobot untuk jawaban “Iris-versicolor” akan bernilai negatif (-) dan bobot jawaban “Iris-setosa” akan bernilai positif (+). Total dari seluruh bobot akan menjadi jawaban akhir. Ilustrasi ini akan disajikan pada Tabel 25.

Tabel 25. Tabel Hasil pengujian kedua algoritma pada data uji

<i>C 4.5</i>	<i>C 4.5 + ADABOOST</i>			
	POHON 1	POHON 2	HASIL	
Iris-setosa	0,3679	-0,4438	-0,0759	Iris-versicolor
Iris-setosa	0,3679	-0,4438	-0,0759	Iris-versicolor
Iris-versicolor	-0,3679	-0,4438	-0,8117	Iris-versicolor
Iris-versicolor	-0,3679	-0,4438	-0,8117	Iris-versicolor
Iris-versicolor	-0,3679	0,4438	0,0759	Iris-setosa
Iris-setosa	0,3679	0,4438	0,8117	Iris-setosa
Iris-versicolor	-0,3679	-0,4438	-0,8117	Iris-versicolor
Iris-setosa	0,3679	-0,4438	-0,0759	Iris-versicolor
Iris-setosa	0,3679	0,4438	0,8117	Iris-setosa
Iris-versicolor	-0,3679	-0,4438	-0,8117	Iris-versicolor
Iris-versicolor	-0,3679	0,4438	0,0759	Iris-setosa
Iris-setosa	0,3679	-0,4438	-0,0759	Iris-versicolor
Iris-setosa	0,3679	0,4438	0,8117	Iris-setosa
Iris-setosa	0,3679	-0,4438	-0,0759	Iris-versicolor
Iris-setosa	0,3679	-0,4438	-0,0759	Iris-versicolor

Berdasarkan tabel 25, pengujian menggunakan *C 4.5* dilakukan dengan menerapkan hukum pada data uji. Pengujian menggunakan *C 4.5* berbasis *Adaboost* dilakukan dengan memperhatikan bobot pohon. Misalnya, pada pohon 1

bobot 0,3679 untuk data uji nomor 1, menghasilkan jawaban “Iris-versicolor”, maka bobot bernilai negatif. Jika menghasilkan jawaban “Iris-setosa” maka bobot bernilai positif. Hal ini juga berlaku pada pengujian oleh pohon 2. Proses berikutnya adalah menjumlahkan hasil dari Pohon 1 dengan Pohon 2. Jika hasilnya negatif, maka jawaban akhir adalah “Iris-versicolor”, jika positif maka “Iris-setosa”.

Proses selanjutnya adalah *Confusion Matriks*. Tabel 3.36 menunjukkan hasil akurasi masing-masing algoritma menggunakan *Confusion Matriks*.

Tabel 26. Tabel Hasil akurasi kedua algoritma

AKURASI					
C 4.5			C 4.5 + ADABOOST		
	DATA TEST			DATA TEST	
PREDIKSI	Iris-setosa	Iris-versicolor	PREDIKSI	Iris-setosa	Iris-versicolor
Iris-setosa	4	5	Iris-setosa	4	1
Iris-versicolor	4	2	Iris-versicolor	4	6

AKURASI	40	AKURASI	66,66 667
ERROR	60	ERROR	33,33 333

Berdasarkan Tabel 26, akurasi *C 4.5* berbasis *Adaboost* jauh lebih tinggi dari pada algoritma *C 4.5*. Algoritma *C 4.5* menghasilkan akurasi 40% sedangkan *C 4.5* berbasis *Adaboost* 66,667%.

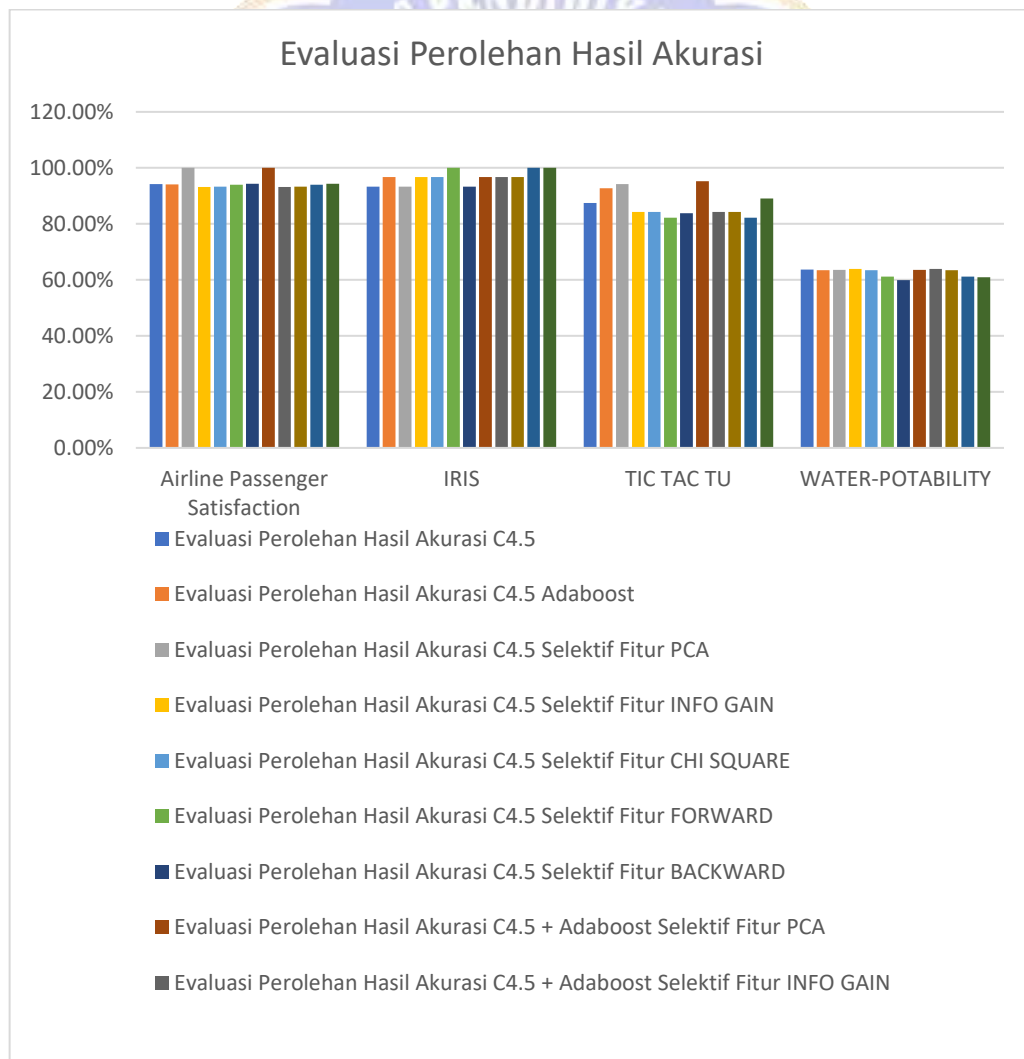
Lampiran 3. Hasil Perhitungan Dengan Sistem

Tabel 27. Tabel Evaluasi Perolehan Hasil Akurasi

DATASET	EVALUASI PEROLEHAN HASIL AKURASI											
	C4.5	C4.5 BERBASIS ADABOOST	C4.5 SELEKTIF FITUR					C4.5 + ADABOOST SELEKTIF FITUR				
			PCA	Info Gain	Chi Square	Forward	Backward	PCA	Info Gain	Chi Square	Forward	Backward
<i>Airline Passenger Satisfaction</i>	94.25 %	94.05%	100.00 %	93.21 %	93.33 %	93.96%	94.34 %	100.00 %	93.21 %	93.33 %	93.96%	94.34%
<i>IRIS</i>	93.33 %	96.67%	93.33%	96.67 %	96.67 %	100.00 %	93.33 %	96.67%	96.67 %	96.67 %	100.00 %	100.00 %
<i>TIC TAC TU</i>	87.43 %	92.67%	94.24%	84.29 %	84.29 %	82.20%	83.77 %	95.29%	84.29 %	84.29 %	82.20%	89.01%
<i>WATER-POTABILITY</i>	63.68 %	63.41%	63.57%	63.87 %	63.41 %	61.13%	59.91 %	63.57%	63.87 %	63.41 %	61.13%	60.98%

*Noted: persentase berwarna biru menandakan *value* naik dibandingkan *value* persentase C4.5 dan C4.5 berbasis AdaBoost

Berdasarkan tabel 27 dapat kita ketahui bahwa dari ke empat dataset yang telah diujikan menunjukkan bahwa algoritma C4.5 berbasis AdaBoost tidak selalu menunjukkan performa terbaik dibandingkan jika algoritma C4.5 dipadankan dengan selektif fitur. Dari *value* persentase juga memperlihatkan beberapa *anomaly* perubahan peningkatan *value* persentase. Peneliti memaknai hal ini sebagai bentuk realitas keberagaman data yang dilakukan uji berupa 4 dataset yang memiliki kriteria berbeda. Selain hal tersebut, diketahui ketidakteraturan perubahan akurasi algoritma C4.5 dan C4.5 berbasis AdaBoost yang dipadankan dengan metode selektif fitur disebabkan oleh kesesuaian dataset dengan kriteria yang paling optimal yang dapat di proses oleh masing-masing selektif fitur. Berikut ini merupakan diagram dari perolehan hasil akurasi kategori semua pengujian yang terlihat pada gambar 9 dibawah ini.



Gambar 9. Gambar Evaluasi Perolehan Hasil Akurasi

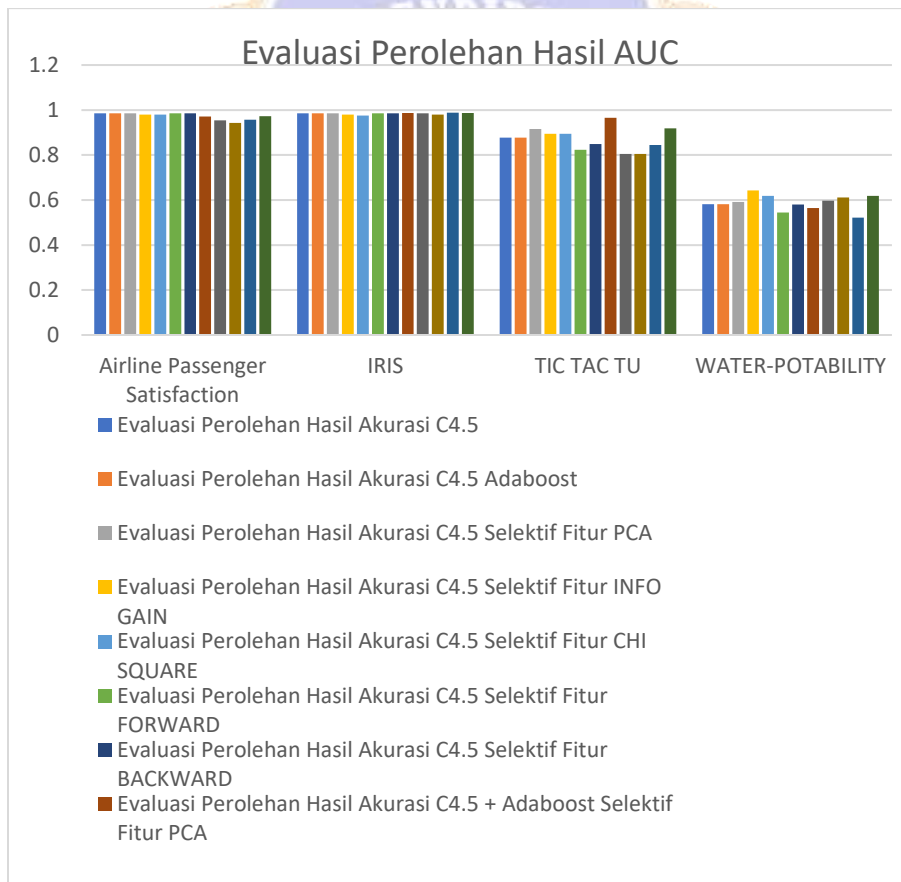
Kemudian untuk perolehan hasil keseluruhan nilai AUC pengujian satu, pengujian dua, pengujian tiga dan pengujian empat dapat kita lihat pada Tabel 28 sebagai berikut.

Tabel 28. Tabel Evaluasi Perolehan Hasil AUC

DATASET	Evaluasi Perolehan Hasil AUC											
	C4.5	C4.5 AdaBoost	C4.5 Selektif Fitur					C4.5 + AdaBoost Selektif Fitur				
			PCA	Info Gain	Chi Square	Forward	Backward	PCA	Info Gain	Chi Square	Forward	Backward
<i>Airline Passenger Satisfaction</i>	0.985	0.985	0.986	0.980	0.98	0.986	0.985	0.971	0.954	0.943	0.957	0.972
<i>IRIS</i>	0.985	0.985	0.985	0.980	0.975	0.985	0.985	0.987	0.985	0.98	0.988	0.987
<i>TIC TAC TU</i>	0.877	0.877	0.916	0.894	0.894	0.824	0.849	0.965	0.805	0.805	0.844	0.919
<i>WATER- POTABILITY</i>	0.582	0.582	0.591	0.642	0.619	0.545	0.58	0.565	0.597	0.611	0.522	0.618

*Noted: persentase berwarna biru menandakan *value* naik dibandingkan *value* pada C4.5 dan C4.5 berbasis AdaBoost

Berdasarkan tabel diatas, dapat dijelaskan bahwa keberhasilan positif(sangat kuat) pada algoritma C4.5 dan C4.5 berbasis AdaBoost dalam menuntaskan prediksi sangat baik pada dataset *Airline Passenger Satisfaction*, *IRIS* dan *TIC TAC TU*. Sedangkan pada dataset *Water-Potability* memiliki katagori keberhasilan prediksi yang kurang baik(sangat lemah). Dataset *Water-Potability* memiliki jumlah *missing value* cukup banyak. Dalam penelitian ini peneliti menerapkan modul *Replace Missing Value* saat eksekusi preprocessing data, hal ini menyebabkan pengujian menghasilkan tingkat keberhasilan yang kurang baik(sangat lemah). Dengan menerapkan selektif fitur dalam pengujian dianggap mampu menghasilkan tingkat akurasi AUC yang sangat baik tanpa melakukan metode *Boosting*. Berikut ini merupakan diagram dari perolehan hasil AUC dari kategori semua pengujian yang terlihat pada gambar 10 dibawah ini.



Gambar 10. Gambar Evaluasi Perolehan Hasil AUC

RIWAYAT HIDUP



Gede Ajus Setiawan lahir di Denpasar pada tahun 1995. Penulis lahir dari pasangan suami istri Bapak Alm. Made Astika dan Ibu Ni Wayan Kerti Artini. Penulis berkebangsaan Indonesia dan beragama Hindu. Kini penulis beralamat di banjar dinas Lemaya Desa Lemukih, Kecamatan Sawan, Kabupaten Buleleng, Provinsi Bali. Penulis menyelesaikan pendidikan dasar di SD Negeri Sanur pada tahun 2005. Kemudian penulis melanjutkan di SMP Negeri 9 Denpasar dan lulus pada tahun 2010. Pada tahun 2013, penulis lulus dari SMK Negeri 1 Denpasar jurusan Rekayasa Perangkat Lunak dan melanjutkan ke Jurusan S1 Pendidikan Teknik Informatika di Universitas Pendidikan Ganesha. Selanjutnya, dari November 2018 sampai dengan penulisan tesis ini penulis masih terdaftar sebagai mahasiswa S2 Ilmu Komputer di Universitas Pendidikan Ganesha.

