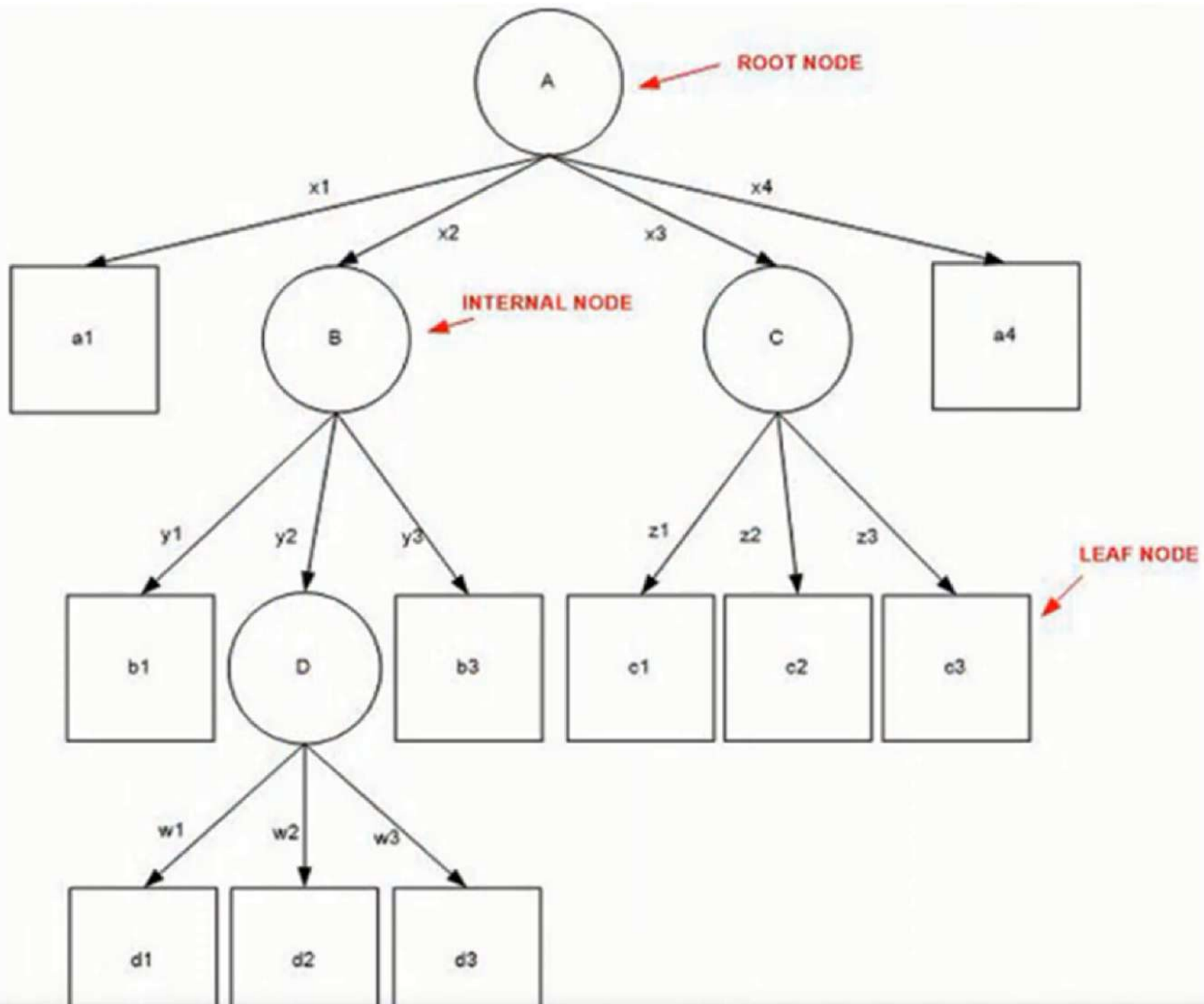


Perhitungan Algoritma ID3

Contoh Procedural Knowledge

Iterative Dichotomiser 3 (ID3)

- Algoritma ID3 merupakan algoritma untuk membangun decision tree atau pohon keputusan
- Ditemukan oleh J. Ross Quinlan (1979), dengan memanfaatkan teori informasi milik Shanon
- Proses dalam decision tree yaitu **mengubah bentuk data (tabel) menjadi model pohon (tree)** kemudian mengubah model pohon tersebut **menjadi aturan (rule)**



Algoritma ID3

1. Menyiapkan dataset
2. Menghitung nilai entropy
3. Menghitung nilai gain
4. Membuat node berdasarkan nilai gain terbesar
5. Ulangi langkah 2 sampai 4 untuk semua node

Entropy

- Entropy adalah keberagaman
- Entropy merupakan parameter untuk mengukur heterogenitas dalam suatu himpunan data

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

- S = himpunan data
- K = banyaknya partisi S
- P_j = Probabilitas yang didapatkan dari “Kasus dibagi Total Kasus”

Information Gain

- Information Gain adalah Perolehan Informasi
- Information Gain merupakan ukuran efektivitas suatu atribut atau variabel dalam mengklasifikasikan data

$$\text{Gain}(A) = \text{entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{entropy}(S_i)$$

- S = ruang (data) sample
- A = Atribut atau variabel
- |S| = jumlah seluruh sample data
- |S_i| = jumlah sample untuk nilai ke i
- Entropy(S_i) = entropy sample yang memiliki nilai i

Menyiapkan Dataset

Kondisi	Temperatur	Kelembapan	Angin	Play Tennis
Cerah	Panas	Tinggi	Lemah	No
Cerah	Panas	Tinggi	Kencang	No
Mendung	Panas	Tinggi	Lemah	Yes
Hujan	Sejuk	Tinggi	Lemah	Yes
Hujan	Dingin	Normal	Lemah	Yes
Hujan	Dingin	Normal	Kencang	No
Mendung	Dingin	Normal	Kencang	Yes
Cerah	Sejuk	Tinggi	Lemah	No
Cerah	Dingin	Normal	Lemah	Yes
Hujan	Sejuk	Normal	Lemah	Yes
Cerah	Sejuk	Normal	Kencang	Yes
Mendung	Sejuk	Tinggi	Kencang	Yes
Mendung	Panas	Normal	Lemah	Yes
Hujan	Sejuk	Tinggi	Kencang	No

Perulangan 1

1. Menghitung Entropy Total Data

- Jumlah Data = 14
- Play Yes = 10
- Play No = 4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Data	Yes	No	Entropy
14	10	4	0,863120569

$$S = \left(-\left(\frac{10}{14}\right) \times \log_2 \left(\frac{10}{14}\right) \right) + \left(-\left(\frac{4}{14}\right) \times \log_2 \left(\frac{4}{14}\right) \right) = 0,863120569$$

2. Menghitung Entropy Per-Kelas

Cuaca	Temperatur	Kelembapan	Angin	Play Tennis
Cerah	Panas	Tinggi	Lemah	No
Cerah	Panas	Tinggi	Kencang	No
Mendung	Panas	Tinggi	Lemah	Yes
Hujan	Sejuk	Tinggi	Lemah	Yes
Hujan	Dingin	Normal	Lemah	Yes
Hujan	Dingin	Normal	Kencang	No
Mendung	Dingin	Normal	Kencang	Yes
Cerah	Sejuk	Tinggi	Lemah	No
Cerah	Dingin	Normal	Lemah	Yes
Hujan	Sejuk	Normal	Lemah	Yes
Cerah	Sejuk	Normal	Kencang	Yes
Mendung	Sejuk	Tinggi	Kencang	Yes
Mendung	Panas	Normal	Lemah	Yes
Hujan	Sejuk	Tinggi	Kencang	No

Cuaca	Jumlah	Yes	No
Cerah	5	2	3
Mendung	4	4	0
Hujan	5	4	1
Total	14	10	4

Temperatur	Jumlah	Yes	No
Panas	4	2	2
Dingin	4	4	0
Sejuk	6	4	2
Total	14	10	4

Kelembapan	Jumlah	Yes	No
Tinggi	7	3	4
Normal	7	7	0
Total	14	10	4

Angin	Jumlah	Yes	No
Kencang	8	6	2
Lemah	6	4	2
Total	14	10	4

2. Menghitung Entropy Cuaca-Cerah

- Jumlah Cerah = 5
- Play Yes = 2
- Play No = 3

Cuaca	Jumlah	Yes	No
Cerah	5	2	3
Mendung	4	4	0
Hujan	5	4	1
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Cerah	Yes	No	Entropy
5	2	3	0,970950594

$$S = \left(-\left(\frac{2}{5}\right) \times \log_2 \left(\frac{2}{5}\right) \right) + \left(-\left(\frac{3}{5}\right) \times \log_2 \left(\frac{3}{5}\right) \right) = 0,970950594$$

2. Menghitung Entropy Cuaca-Mendung

- Jumlah Mendung = 4
- Play Yes = 4
- Play No = 0

Cuaca	Jumlah	Yes	No
Cerah	5	2	3
Mendung	4	4	0
Hujan	5	4	1
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Mendung	Yes	No	Entropy
4	4	0	0

$$S = \left(-\left(\frac{4}{4}\right) \times \log_2 \left(\frac{4}{4}\right) \right) + \left(-\left(\frac{0}{4}\right) \times \log_2 \left(\frac{0}{4}\right) \right) = 0$$

2. Menghitung Entropy Cuaca-Hujan

- Jumlah Hujan = 5
- Play Yes = 4
- Play No = 1

Cuaca	Jumlah	Yes	No
Cerah	5	2	3
Mendung	4	4	0
Hujan	5	4	1
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Hujan	Yes	No	Entropy
5	4	1	0,721928095

$$S = \left(-\left(\frac{4}{5}\right) \times \log_2 \left(\frac{4}{5}\right) \right) + \left(-\left(\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right) \right) = 0,721928095$$

2. Menghitung Entropy Temperatur-Panas

- Jumlah Panas = 4
- Play Yes = 2
- Play No = 2

Temperatur	Jumlah	Yes	No
Panas	4	2	2
Dingin	4	4	0
Sejuk	6	4	2
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Panas	Yes	No	Entropy
4	2	2	1

$$S = \left(-\left(\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) \right) + \left(-\left(\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) \right) = 1$$

2. Menghitung Entropy Temperatur-Dingin

- Jumlah Dingin = 4
- Play Yes = 4
- Play No = 0

Temperatur	Jumlah	Yes	No
Panas	4	2	2
Dingin	4	4	0
Sejuk	6	4	2
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Dingin	Yes	No	Entropy
4	4	0	0

$$S = \left(-\left(\frac{4}{4}\right) \times \log_2 \left(\frac{4}{4}\right) \right) + \left(-\left(\frac{0}{4}\right) \times \log_2 \left(\frac{0}{4}\right) \right) = 0$$

2. Menghitung Entropy Temperatur-Sejuk

- Jumlah Sejuk = 6
- Play Yes = 4
- Play No = 2

Temperatur	Jumlah	Yes	No
Panas	4	2	2
Dingin	4	4	0
Sejuk	6	4	2
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Sejuk	Yes	No	Entropy
6	4	2	0,918295834

$$S = \left(-\left(\frac{4}{6}\right) \times \log_2 \left(\frac{4}{6}\right) \right) + \left(-\left(\frac{2}{6}\right) \times \log_2 \left(\frac{2}{6}\right) \right) = 0,918295834$$

2. Menghitung Entropy Kelembapan-Tinggi

- Jumlah Tinggi = 7
- Play Yes = 3
- Play No = 4

Kelembapan	Jumlah	Yes	No
Tinggi	7	3	4
Normal	7	7	0
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Tinggi	Yes	No	Entropy
7	3	4	0,985228136

$$S = \left(-\left(\frac{3}{7}\right) \times \log_2 \left(\frac{3}{7}\right) \right) + \left(-\left(\frac{4}{7}\right) \times \log_2 \left(\frac{4}{7}\right) \right) = 0,985228136$$

2. Menghitung Entropy Kelembapan-Normal

- Jumlah Normal = 7
- Play Yes = 7
- Play No = 0

Kelembapan	Jumlah	Yes	No
Tinggi	7	3	4
Normal	7	7	0
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Normal	Yes	No	Entropy
7	7	0	0

$$S = \left(-\left(\frac{7}{7}\right) \times \log_2 \left(\frac{7}{7}\right) \right) + \left(-\left(\frac{0}{7}\right) \times \log_2 \left(\frac{0}{7}\right) \right) = 0$$

2. Menghitung Entropy Angin-Kencang

- Jumlah Kencang = 8
- Play Yes = 6
- Play No = 2

Angin	Jumlah	Yes	No
Kencang	8	6	2
Lemah	6	4	2
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Kencang	Yes	No	Entropy
8	6	2	0,918295834

$$S = \left(-\left(\frac{6}{8}\right) \times \log_2 \left(\frac{6}{8}\right) \right) + \left(-\left(\frac{2}{8}\right) \times \log_2 \left(\frac{2}{8}\right) \right) = 0,918295834$$

2. Menghitung Entropy Angin-Lemah

- Jumlah Lemah = 6
- Play Yes = 4
- Play No = 2

Angin	Jumlah	Yes	No
Kencang	8	6	2
Lemah	6	4	2
Total	14	10	4

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Lemah	Yes	No	Entropy
6	4	2	0,811278124

$$S = \left(-\left(\frac{4}{6}\right) \times \log_2 \left(\frac{4}{6}\right) \right) + \left(-\left(\frac{2}{6}\right) \times \log_2 \left(\frac{2}{6}\right) \right) = 0,811278124$$

2. Hasil Nilai Entropy

		Data	Yes	No	Entropy
Total		14	10	4	0,863120569
Cuaca					
	Cerah	5	2	3	0,970950594
	Mendung	4	4	0	0,00
	Hujan	5	4	1	0,721928095
Temperatur					
	Panas	4	2	2	1,00
	Dingin	4	4	0	0,00
	Sejuk	6	4	2	0,918295834
Kelembapan					
	Tinggi	7	3	4	0,985228136
	Normal	7	7	0	0,00
Angin					
	Lemah	8	6	2	0,811278124
	Kencang	6	4	2	0,918295834

3. Menghitung Nilai Gain - Cuaca

$$\text{Gain}(A) = \text{entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{S} \times \text{entropy}(S_i)$$

		Data	Entropy
Total		14	0,863120569
Cuaca			
	Cerah	5	0,970950594
	Mendung	4	0,00
	Hujan	5	0,721928095

Gain Cuaca =

Total Entorpy - ((Jml_Cerah/Data * Entropy Cerah) +
(Jml_Mendung)/Data*Entropy Mendung) + (Jml-Hujan/Data*Entropy Hujan))

Gain Cuaca =

0,863120569 - ((5/14 * 0,970950594) + (4/14 * 0) + (5/14 * 0,721928095))

Gain Cuaca = 0,258521037

3. Menghitung Nilai Gain - Temperatur

$$\text{Gain}(A) = \text{entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{S} \times \text{entropy}(S_i)$$

		Data	Entropy
Total		14	0,863120569
Temperatur			
	Panas	4	1,00
	Dingin	4	0,00
	Sejuk	6	0,918295834

Gain Temperatur =

$$\text{Total Entorpy} - ((\text{Jml_Panas}/\text{Data} * \text{Entropy Panas}) + (\text{Jml_Dingin})/\text{Data} * \text{Entropy Dingin}) + (\text{Jml-Sejuk}/\text{Data} * \text{Entropy Sejuk}))$$

Gain Temperatur =

$$0,863120569 - ((4/14 * 1) + (4/14 * 0) + (6/14 * 0,918295834))$$

$$\text{Gain Temperatur} = 0,183850925$$

3. Menghitung Nilai Gain - Kelembapan

$$\text{Gain}(A) = \text{entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{S} \times \text{entropy}(S_i)$$

		Data	Entropy
Total		14	0,863120569
Kelembapan			
	Tinggi	7	0,985228136
	Normal	7	0,00

Gain Kelembapan =

Total Entorpy - ((Jml_Tinggi/Data * Entropy Tinggi) +
(Jml_Normal)/Data*Entropy Normal))

Gain Kelembapan =

0,863120569 - ((7/14 * 0,985228136) + (7/14 * 0))

Gain Kelembapan = 0,370506501

3. Menghitung Nilai Gain - Angin

$$\text{Gain}(A) = \text{entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{S} \times \text{entropy}(S_i)$$

		Data	Entropy
Total		14	0,863120569
Angin			
	Lemah	8	0,811278124
	Kencang	6	0,918295834

Gain Angin =

Total Entorpy - ((Jml_Lemah/Data * Entropy Lemah) +
(Jml_Kencang)/Data*Entropy Kencang))

Gain Angin =

0,863120569 - ((8/14 * 0,811278124) + (6/14 * 0,918295834))

Gain Angin = 0,005977711

3. Hasil Perhitungan Gain

		Data	Yes	No	Entropy	Gain
Total		14	10	4	0,863120569	
Cuaca						0,258521037
	Cerah	5	2	3	0,970950594	
	Mendung	4	4	0	0,00	
	Hujan	5	4	1	0,721928095	
Temperatur						0,183850925
	Panas	4	2	2	1,00	
	Dingin	4	4	0	0,00	
	Sejuk	6	4	2	0,918295834	
Kelembapan						0,370506501
	Tinggi	7	3	4	0,985228136	
	Normal	7	7	0	0,00	
Angin						0,005977771
	Lemah	8	6	2	0,811278124	
	Kencang	6	4	2	0,918295834	

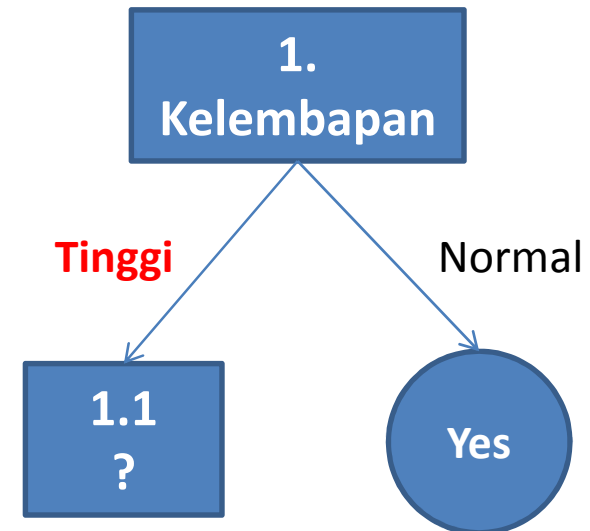
4. Membuat Node Cabang Berdasarkan Gain Yang Terbesar

		Data	Yes	No	Entropy	Gain
Total		14	10	4	0,863120569	
Cuaca						0,258521037
	Cerah	5	2	3	0,970950594	
	Mendung	4	4	0	0,00	
	Hujan	5	4	1	0,721928095	
Temperatur						0,183850925
	Panas	4	2	2	1,00	
	Dingin	4	4	0	0,00	
	Sejuk	6	4	2	0,918295834	
Kelembapan						0,370506501
	Tinggi	7	3	4	0,985228136	
	Normal	7	7	0	0,00	
Angin						0,005977771
	Lemah	8	6	2	0,811278124	
	Kencang	6	4	2	0,918295834	

4. Membuat Node Cabang Berdasarkan Gain Yang Terbesar

		Data	Yes	No	Entropy	Gain
Kelembapan						0,370506501
	Tinggi	7	3	4	0,985228136	
	Normal	7	7	0	0,00	

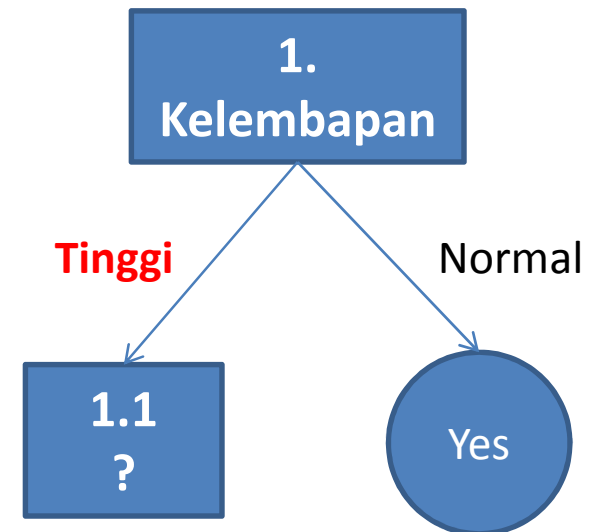
- Nilai gain terbesar adalah **Kelembapan**
- Kelembapan menjadi **node awal / root node**
- Kelembapan memiliki 2 Variabel = **Tinggi dan Normal**
- **Normal** memiliki nilai entropy =0, maka tidak memiliki cabang. Dan "**Yes = 7, No=0**"
- **Tinggi** memiliki nilai entropy >0, maka akan mempunyai cabang, dimana akan menjadi node awal untuk perulangan berikutnya



4. Membuat Node Cabang Berdasarkan Gain Yang Terbesar

- Berdasarkan pembentukan pohon keputusan untuk node 1 (root node), maka dilakukan analisis lebih lanjut untuk menentukan node awal berikutnya
- Dengan cara memfilter data yang memiliki **kelembapan = tinggi**

No	Cuaca	Temperat	Kelembapan	Angin	Play Tennis
1	Cerah	Panas	Tinggi	Lemah	No
2	Cerah	Panas	Tinggi	Kencang	No
3	Mendung	Panas	Tinggi	Lemah	Yes
4	Hujan	Sejuk	Tinggi	Lemah	Yes
5	Cerah	Sejuk	Tinggi	Lemah	No
6	Mendung	Sejuk	Tinggi	Kencang	Yes
7	Hujan	Sejuk	Tinggi	Kencang	No



Perulangan 2

Menyiapkan Dataset

No	Cuaca	Temperat	Kelembapan	Angin	Play Tennis
1	Cerah	Panas	Tinggi	Lemah	No
2	Cerah	Panas	Tinggi	Kencang	No
3	Mendung	Panas	Tinggi	Lemah	Yes
4	Hujan	Sejuk	Tinggi	Lemah	Yes
5	Cerah	Sejuk	Tinggi	Lemah	No
6	Mendung	Sejuk	Tinggi	Kencang	Yes
7	Hujan	Sejuk	Tinggi	Kencang	No

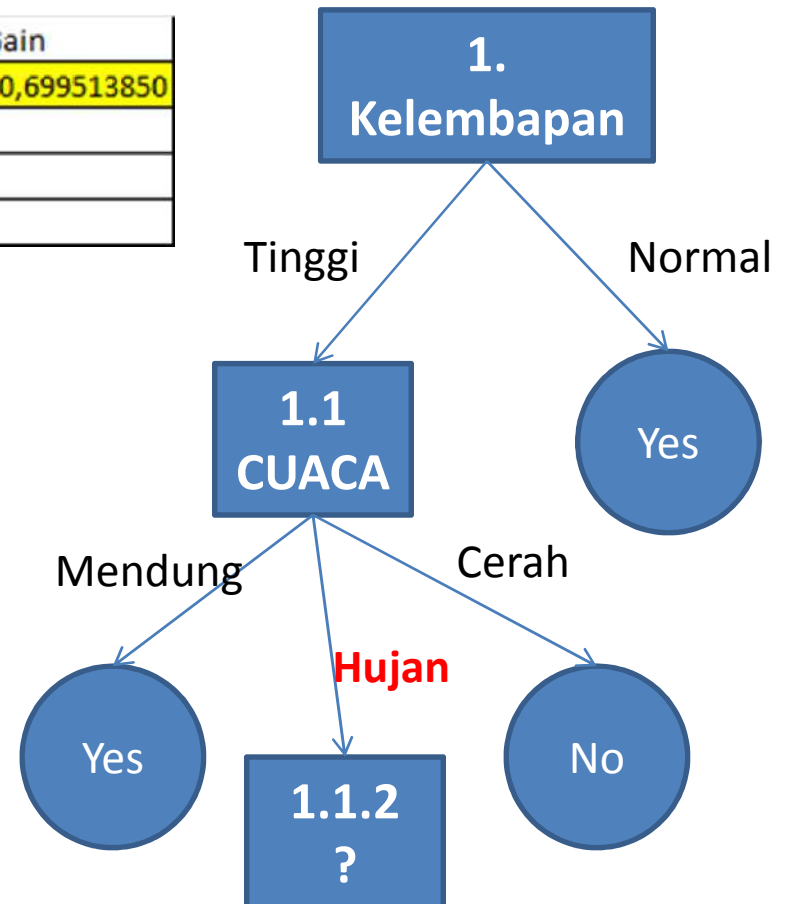
Hasil Perulangan 2

		Data	Yes	No	Entropy	Gain
Total		7	3	4	0,985228136	
Cuaca						0,699513850
	Cerah	3	0	3	0	
	Mendung	2	2	0	0	
	Hujan	2	1	1	1	
Temperatur						0,020244207
	Panas	3	2	1	0,918295834	
	Sejuk	4	2	2	1	
	Dingin	0	0	0	0	
Angin						0,020244207
	Lemah	4	2	2	1	
	Kencang	3	1	2	0,918295834	

Membuat Node Cabang Berdasarkan Gain Yang Terbesar

		Data	Yes	No	Entropy	Gain
Cuaca						0,699513850
	Cerah	3	0	3	0	
	Mendung	2	2	0	0	
	Hujan	2	1	1	1	

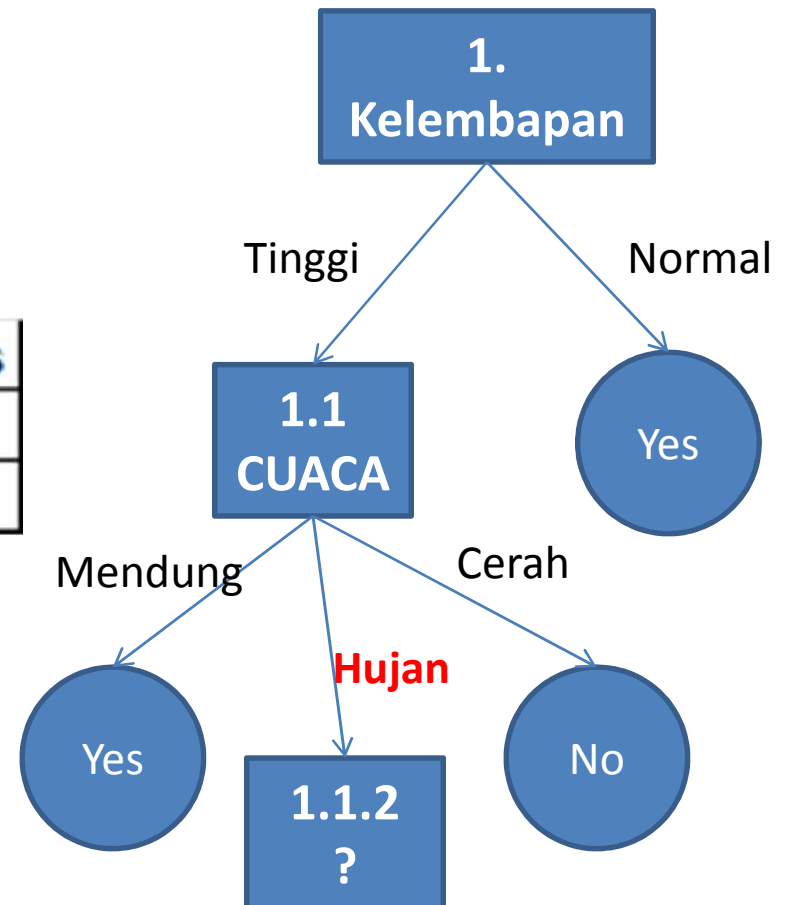
- Nilai gain terbesar adalah **CUACA**
- **CUACA** menjadi **node awal**
- Cuaca memiliki 3 Variabel = **Cerah, Mendung, Hujan**
- **Cerah, Mendung** memiliki nilai entropy = 0, maka tidak memiliki cabang
- Cerah memiliki "**No=3**", yes=0
- Mendung memiliki "**Yes=2**", no = 0
- **Hujan** memiliki nilai entropy > 0, maka akan mempunyai cabang, dimana akan menjadi node awal untuk perulangan berikutnya



Membuat Node Cabang Berdasarkan Gain Yang Terbesar

- Berdasarkan pembentukan pohon keputusan untuk node 1.1, maka dilakukan analisis lebih lanjut untuk menentukan node awal berikutnya
- Dengan cara memfilter data yang memiliki **cuaca = hujan**

No	Cuaca	Temperatur	Angin	Play Tennis
1	Hujan	Sejuk	Lemah	Yes
2	Hujan	Sejuk	Kencang	No



Perulangan 3

Menyiapkan Dataset

No	Cuaca	Temperatur	Angin	Play Tennis
1	Hujan	Sejuk	Lemah	Yes
2	Hujan	Sejuk	Kencang	No

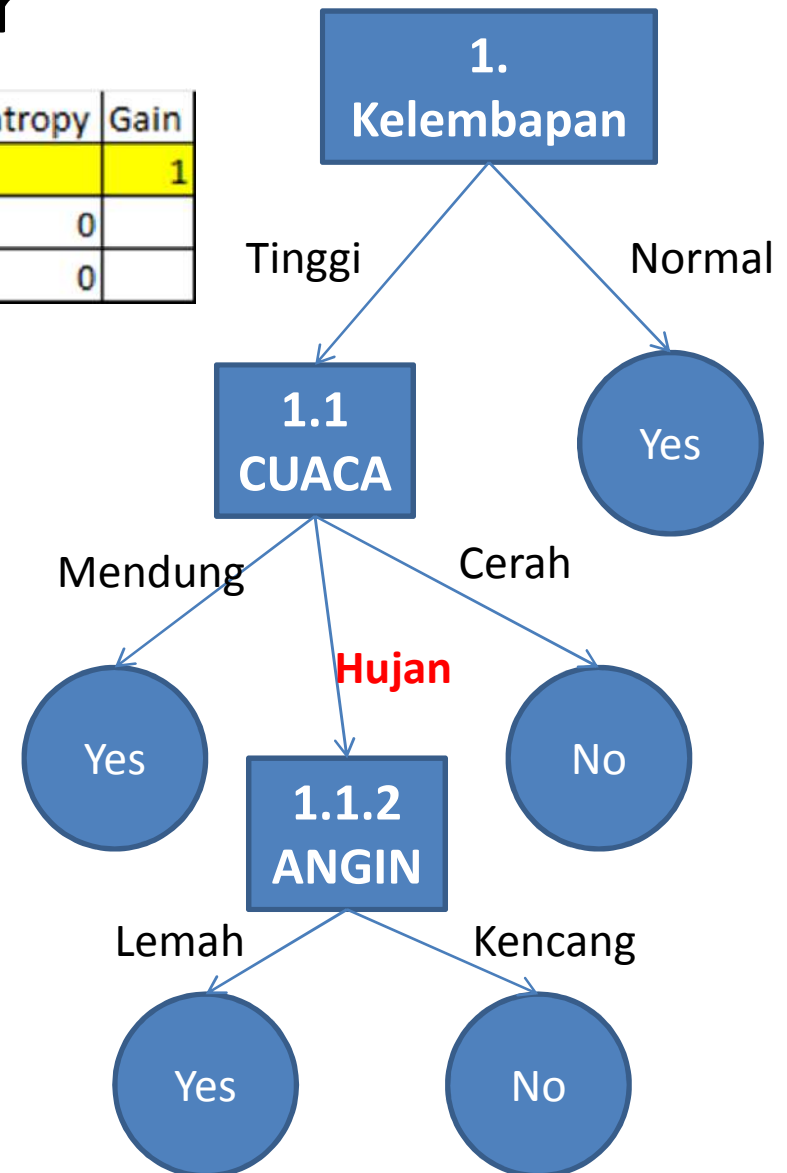
Hasil Perulangan 3

		Data	Yes	No	Entropy	Gain
Total		2	1	1	1	
Temperatur						0
	Panas	0	0	0	0	
	Sejuk	2	1	1	1	
	Dingin	0	0	0	0	
Angin						1
	Lemah	1	1	0	0	
	Kencang	1	0	1	0	

Membuat Node Cabang Berdasarkan Gain Yang Terbesar

		Data	Yes	No	Entropy	Gain
Angin						1
	Lemah	1	1	0	0	
	Kencang	1	0	1	0	

- Nilai gain terbesar adalah **ANGIN**
- **ANGIN** menjadi **node awal**
- Cuaca memiliki 2 Variabel = **lemah, Kencang**
- **Lemah, Kencang** memiliki nilai entropy =0, maka tidak memiliki cabang
- Lemah memiliki "**Yes=1**", no=0
- Kencang memiliki "**No=1**", yes = 0



Hasil Akhir Pohon Keputusan

Dari Pohon Keputusan tersebut maka dapat diperoleh aturan:

1. Jika Kelembapan Normal, maka dapat bermain tenis
2. Jika Kelembapan Tinggi, Cek Cuaca:
 1. Jika Cuaca mendung, maka dapat bermain tenis
 2. Jika Cuaca Cerah, maka tidak dapat bermain tenis
3. Jika Cuaca Hujan, Cek Angin:
 1. Jika Angin Kencang, maka tidak dapat bermain tenis
 2. Jika Angin Lemah, maka dapat bermain tenis

IF Kelembapan=Normal Then Yes

IF Kelembapan=Tinggi ^ Cuaca=Cerah Then No

IF Kelembapan=Tinggi ^ Cuaca=Mendung Then Yes

IF Kelembapan=Tinggi ^ Cuaca=Hujan ^ Angin=Lemah Then Yes

IF Kelembapan=Tinggi ^ Cuaca=Hujan ^ Angin=Kencang Then No

