Decision Tree Induction

Contoh 1:

No	Mahasiswa	Status
1	Aktif	Positif
2	Aktif	Negatif
3	Tidak Aktif	Negatif
4	Tidak Aktif	Negatif
5	Tidak Aktif	Positif
6	Aktif	Positif
7	Tidak Aktif	Negatif
8	Tidak Aktif	Negatif
9	Tidak Aktif	Negatif
10	Aktif	Negatif

$P_+ = 3/10$

$P_- = 7/10$

$$Info(D) = -\sum_{i=1}^{m} p_i \, log_2(p_i)$$

$$Info(D) = \left(-\frac{3}{10} x \, Log_2\left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} x \, Log_2\left(\frac{7}{10}\right)\right) \approx 0.88$$

Contoh 2:

No	Nilai	Bonus
1	Excellent	Normal
2	Excellent	Normal
3	Excellent	Normal
4	Excellent	High
5	Good	Normal
6	Good	Normal
7	Good	Normal
8	Good	Normal
9	Good	High
10	Good	High
11	Poor	High
12	Poor	High
13	Poor	High
14	Poor	High

1. Hitung entropi dari target attribute

$$Info(D) = \left(-\frac{7}{14} x Log_2\left(\frac{7}{14}\right)\right) + \left(-\frac{7}{14} x Log_2\left(\frac{7}{14}\right)\right) = 1$$

2. Hitung entropi dari tiap kelas pada predictor attribute

$$Info_{Nilai}(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

- $Info(D_{Excellent}) = \left(-\frac{3}{4} x Log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} x Log_2\left(\frac{1}{4}\right)\right) = 0{,}811$
- $Info(D_{Good}) = \left(-\frac{4}{6} x Log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} x Log_2\left(\frac{2}{6}\right)\right) = 0{,}918$
- $Info(D_{Poor}) = \left(-0 x Log_2(0)\right) + \left(-\frac{4}{4} x Log_2\left(\frac{4}{4}\right)\right) = 0$

$$Info_{Nilai}(D) = \left(\left(\frac{4}{14} x\ 0{,}811\right) + \left(\frac{6}{14} x\ 0{,}918\right) + \left(\frac{4}{14} x\ 0\right)\right) = 0{,}625$$

3. Hitung Information Gain setiap Attribute prediktor

$Gain(Nilai) = Info(D) - Info_{Nilai}(D)$
$= 1 - 0{,}625$
$= 0{,}375$

Contoh 3:

ID	Outlook	Temperature	Humidity	Wind	Decision
0	Sunny	Hot	High	Weak	No
1	Sunny	Hot	High	Strong	No
2	Overcast	Hot	High	Weak	Yes
3	Rain	Mild	High	Weak	Yes
4	Rain	Cool	Normal	Weak	Yes
5	Rain	Cool	Normal	Strong	Yes
6	Overcast	Cool	Normal	Strong	Yes
7	Sunny	Mild	High	Weak	No
8	Sunny	Cool	Normal	Weak	Yes
9	Rain	Mild	Normal	Weak	Yes
10	Sunny	Mild	Normal	Strong	Yes
11	Overcast	Mild	High	Strong	Yes
12	Overcast	Hot	Normal	Weak	Yes
13	Rain	Mild	High	Strong	No

Langkah 1: Tentukan Root

1. Hitung entropi dari Attribute target

$$Info(D) = \left(-\frac{4}{14} x Log_2\left(\frac{4}{14}\right)\right) + \left(-\frac{10}{14} x Log_2\left(\frac{10}{14}\right)\right) \approx 0{,}863$$

2. Hitung entropi dari tiap kelas pada Attribute prediktor

A. **Outlook**

- $Info(D_{Overcast}) = \left(-\frac{0}{4} x Log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} x Log_2\left(\frac{4}{4}\right)\right) = 0$
- $Info(D_{Rain}) = \left(-\frac{1}{5} x Log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} x Log_2\left(\frac{4}{5}\right)\right) = 0{,}722$
- $Info(D_{Sunny}) = \left(-\frac{3}{5} x Log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} x Log_2\left(\frac{2}{5}\right)\right) = 0{,}971$

$$Info_{Outlook}(D) = \left(\left(\frac{4}{14} x\ 0\right) + \left(\frac{5}{14} x\ 0{,}722\right) + \left(\frac{5}{14} x\ 0{,}971\right)\right) = 0{,}605$$

B. **Temperature**

- $Info(D_{Cool}) = \left(-\frac{0}{4}xLog_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4}xLog_2\left(\frac{4}{4}\right)\right) = 0$
- $Info(D_{Hot}) = \left(-\frac{2}{4}xLog_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4}xLog_2\left(\frac{2}{4}\right)\right) = 1$
- $Info(D_{Mild}) = \left(-\frac{2}{6}xLog_2\left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6}xLog_2\left(\frac{4}{6}\right)\right) = 0{,}918$

$$Info_{Temperature}(D) = \left(\left(\frac{4}{14}x\ 0\right) + \left(\frac{4}{14}x\ 1\right) + \left(\frac{6}{14}x\ 0{,}918\right)\right) = 0{,}679$$

C. **Humidity**

$$Info_{Humidity}(D) = 0{,}493$$

D. **Wind**

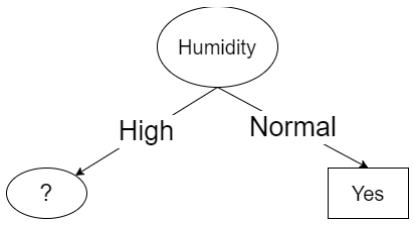
$$Info_{Wind}(D) = 0{,}857$$

3. Hitung Information Gain setiap Attribute prediktor

$Gain(Outlook) = Info(D) - Info_{Outlook}(D)$
$= 0{,}863 - 0{,}0679$
$= 0{,}259$
$Gain(Temperature) = Info(D) - Info_{Temperature}(D)$
$= 0{,}863 - 0{,}679$
$= 0{,}184$
$Gain(Humidity) = Info(D) - Info_{Humidity}(D)$
$= 0{,}863 - 0{,}493$
$= 0{,}371$
$Gain(Wind) = Info(D) - Info_{Wind}(D)$
$= 0{,}863 - 0{,}857$
$= 0{,}006$

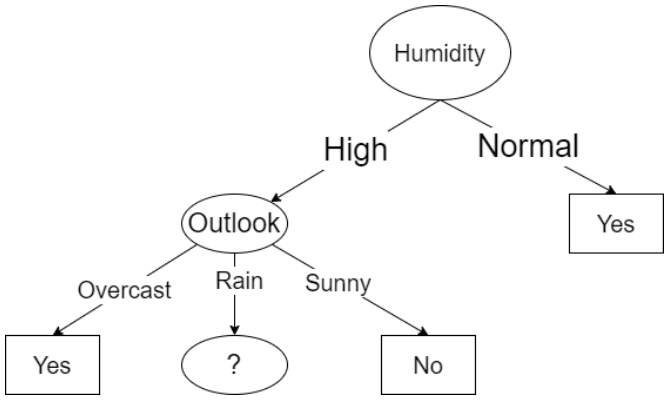
Attribute	Decision			Entropy	Information Gain
	Yes	No	Total		
Outlook					
• Overcast	4	0	4	0	0,256
• Rain	4	1	5	0,722	
• Sunny	2	3	5	0,971	
Temperature					
• Cool	0	4	4	0	0,184
• Hot	2	2	4	1	
• Mild	2	4	6	0,918	
Humidity					
• High	4	3	7	0,985	0,371
• Normal	7	0	7	0	
Wind					
• Weak	2	6	8	0,985	0,006
• Strong	4	2	6	0,918	

ID	Outlook	Temperature	Humidity	Wind	Decision
0	Sunny	Hot	High	Weak	No
1	Sunny	Hot	High	Strong	No
2	Overcast	Hot	High	Weak	Yes
3	Rain	Mild	High	Weak	Yes
4	Rain	Cool	Normal	Weak	Yes
5	Rain	Cool	Normal	Strong	Yes
6	Overcast	Cool	Normal	Strong	Yes
7	Sunny	Mild	High	Weak	No
8	Sunny	Cool	Normal	Weak	Yes
9	Rain	Mild	Normal	Weak	Yes
10	Sunny	Mild	Normal	Strong	Yes
11	Overcast	Mild	High	Strong	Yes
12	Overcast	Hot	Normal	Weak	Yes
13	Rain	Mild	High	Strong	No



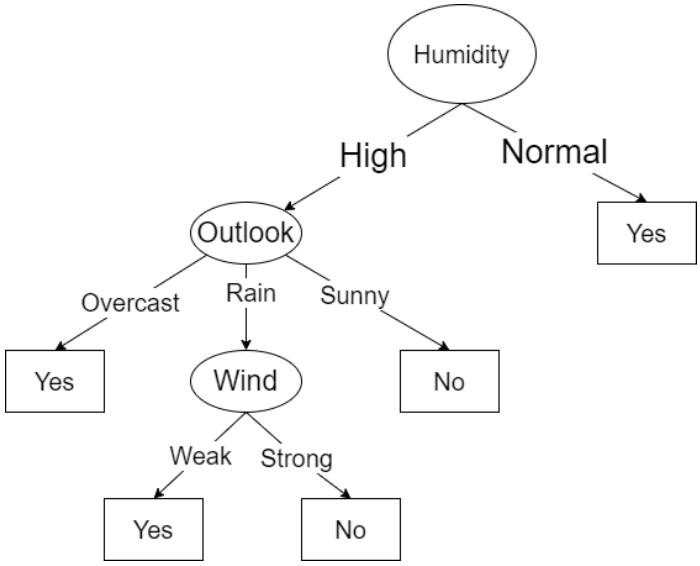
ID	Outlook	Temperature	Humidity	Wind	Decision
0	Sunny	Hot	High	Weak	No
1	Sunny	Hot	High	Strong	No
2	Overcast	Hot	High	Weak	Yes
3	Rain	Mild	High	Weak	Yes
7	Sunny	Mild	High	Weak	No
11	Overcast	Mild	High	Strong	Yes
13	Rain	Mild	High	Strong	No

Attribute	Decision			Entropy	Information Gain
	Yes	No	Total		
Humidity|High	4	3	7	0,985	
Outlook					
• Overcast	2	0	2	0	0,6995
• Rain	1	1	2	1	
• Sunny	0	3	3	0	
Temperature					
• Cool	0	0	0	0	0,0203
• Hot	1	2	3	0,918	
• Mild	2	2	4	1	
Wind					
• Weak	2	2	4	1	0,0203
• Strong	1	2	3	0,918	



ID	Outlook	Temperature	Humidity	Wind	Decision
3	Rain	Mild	High	Weak	Yes
13	Rain	Mild	High	Strong	No

Attribute	Decision			Entropy	Information Gain
	Yes	No	Total		
Humidity|High and Outlook| Rain	1	1	2	1	
Temperature					
• Cool	0	0	0	0	0
• Hot	0	0	0	0	
• Mild	1	1	2	1	
Wind					
• Weak	1	0	1	0	1
• Strong	0	1	1	0	



Aturan If-Else dari hasil Decision Tree

- If Humidity = Normal, then Yes
- If Humidity = High and Outlook = Sunny, then No
- If Humidity = High and Outlook = Overcast, then Yes
- If Humidity = High and Outlook = Rain and Wind = Strong then No
- If Humidity = High and Outlook = Rain and Wind = Weak then Yes