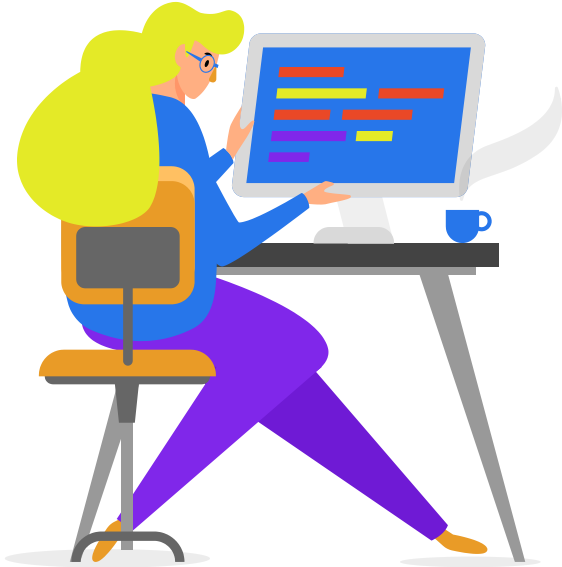


# Kecerdasan Buatan Pertemuan 8

IFC22J3  
Informatika

# Pokok Bahasan



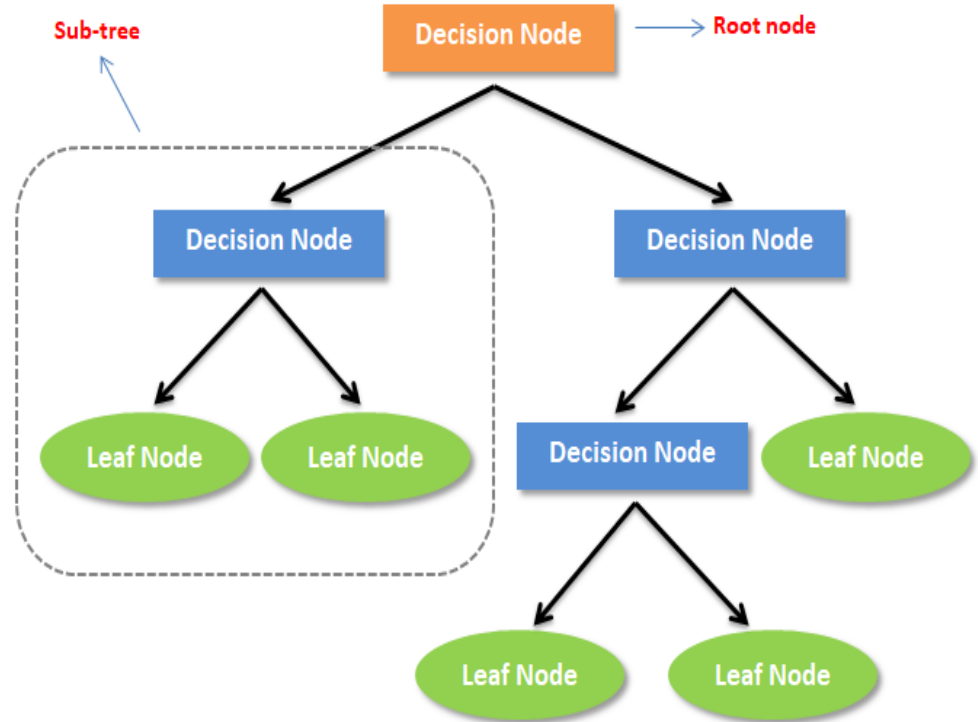
**01** Decision tree

**02** Struktur

**03** Entropy, Gain

# Decision tree

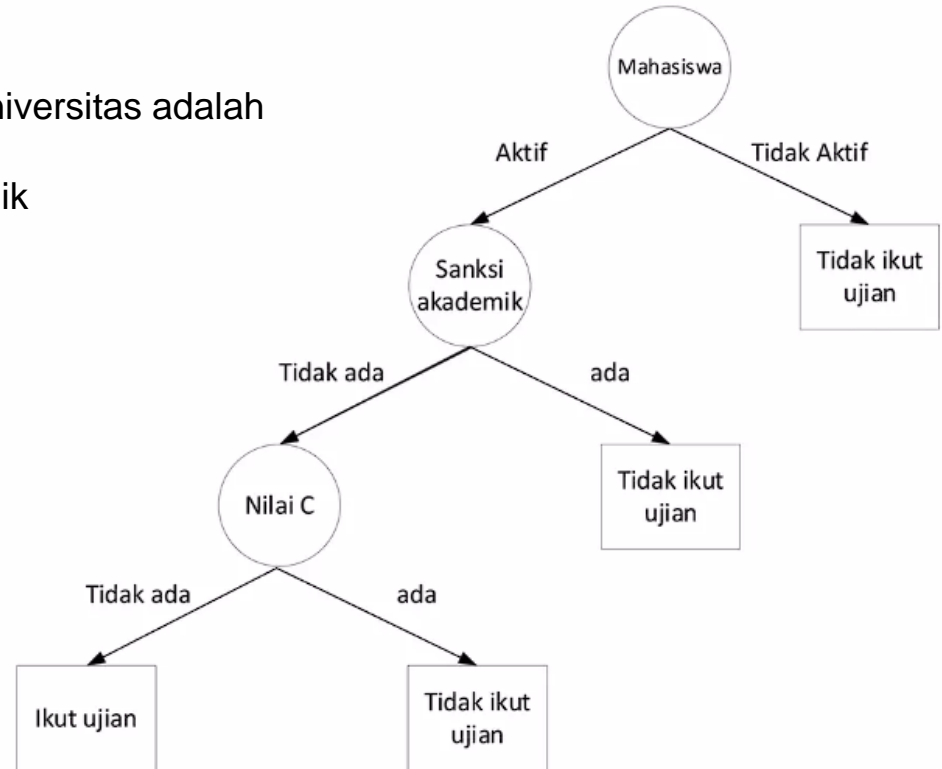
Decision tree adalah sebuah metode analisis dan pemodelan yang menggunakan struktur pohon untuk membuat keputusan atau melakukan prediksi. Struktur pohon ini memiliki simpul-simpul yang mewakili tes atau pemisahan berdasarkan atribut-atribut data, dan cabang-cabang yang mewakili kemungkinan hasil atau konsekuensi dari tes tersebut.



# Contoh

Misalkan syarat mengikuti ujian pada sebuah universitas adalah

- Terdaftar sebagai mahasiswa aktif
- Tidak sedang mendapatkan sanksi akademik
- Tidak ada nilai C



# Dasar decision tree

1. Mencari ukuran varians / entropy dari berbagai atribut
2. Mendapatkan akar sebagai atribut yang memiliki informasi terbanyak dalam data set berdasarkan ukuran entropy
3. Membuat cabang untuk atribut yang memiliki peringkat dibawah akar, semakin kecil informasi yang dimiliki, atribut akan semakin dbawah
4. Ujung dari DT adalah daun, di tingkat ini perhitungan akan berhenti dilakukan karena mencapai akhir pohon
5. Tahap akhir akan terbentuk pohon dan menghasilkan aturan IF – Then yang dapat digunakan untuk menentukan permintaan sesuai dengan aturan

## **Kenapa root node penting?**

1. Jika menggunakan sembarang atribut sebagai root, maka aturan yang di hasilkan tidak akan efektif
2. Mencari atribut sebagai root untuk mengetahui seberapa pengaruhnya atribut tersebut pada kelas target
3. Tanpa root tidak akan bisa membuat struktur pohon yang benar

## Menentukan root

- Menghitung varians dari setiap kelas pada atribut dan total target
- Hitung varians dengan entropy
- Cari bobot dari setiap atribut
- Hitung entropy total dikurangi rata-tata bobot => information gain
- Atribut dengan information gaib tertinggi layak digunakan sebagai root

# Entropy

Ukuran ketidakteraturan (measure of disorder)

Rumusnya:

$$E(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Keterangan:

- $S$  = Himpunan kasus
- $n$  = Jumlah anggota dari  $S$
- $p_i$  = proporsi dari  $S_i$  terhadap  $S$



## Contoh entropy

Ada sebuah data set yang terbagi menjadi 2 kelas positif dan negatif, dengan total baris sebanyak 10 yang terbagi atas 3 baris positif dan 7 baris negatif

$$E(S) = \sum_{i=1}^n -p_i \log_2 p_i$$
$$E(S) = \left( -\frac{3}{10} \times \log_2 \left( \frac{3}{10} \right) \right) + \left( -\frac{7}{10} \times \log_2 \left( \frac{7}{10} \right) \right) \approx 0.88$$

# Information Gain

Merupakan sebuah teknik untuk mengurangi ketidakaturan entropy sehingga mendapatkan ukuran informasi dari hasilnya, semakin tinggi hasil berarti semakin tinggi pula informasi yang didapatkan

$$IG(Y, X) = E(Y) - E(Y|X)$$

$E(Y)$	Entropy dari atribut
$E(Y X)$	Entropy rata-rata dari atribut prediktor terhadap atribut target
$X$	Atribut Prediktor
$Y$	Atribut Target

# Contoh

No	Pelanggaran	Bonus
1	Tinggi	Normal
2	Tinggi	Normal
3	Tinggi	Normal
4	Tinggi	Tinggi
5	Sedang	Normal
6	Sedang	Normal
7	Sedang	Normal
8	Sedang	Normal
9	Sedang	Tinggi
10	Sedang	Tinggi
11	Rendah	Tinggi
12	Rendah	Tinggi
13	Rendah	Tinggi
14	Rendah	Tinggi

Menentukan entropy dari target bonus

$$E(\text{Bonus}) = \left(-\frac{7}{14} \times \log_2\left(\frac{7}{14}\right)\right) + \left(-\frac{7}{14} \times \log_2\left(\frac{7}{14}\right)\right) = 1$$

Menentukan entropy dari tiap kelas

$$E(\text{Bonus}|\text{Pelanggaran} = \text{Excellent}) = \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) \approx 0.811$$

$$E(\text{Bonus}|\text{Pelanggaran} = \text{Good}) = \left(-\frac{4}{6} \times \log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} \times \log_2\left(\frac{2}{6}\right)\right) \approx 0.918$$

$$E(\text{Bonus}|\text{Pelanggaran} = \text{Poor}) = (-0 \times \log_2(0)) + \left(-\frac{4}{4} \times \log_2\left(\frac{4}{4}\right)\right) = 0$$

Menentukan rata-rata bobot atribut pelanggaran

$$E(\text{Bonus}|\text{Pelanggaran}) = \left(\left(\frac{4}{14} \times 0.811\right) + \left(\frac{6}{14} \times 0.918\right) + \left(\frac{4}{14} \times 0\right)\right) = 0.625$$

## Penyelesaian

$$IG(Y, X) = E(Y) - E(Y|X)$$

$$\begin{aligned} IG(\text{Bonus}, \text{Pelanggaran}) &= E(\text{Bonus}) - E(\text{Bonus}|\text{Pelanggaran}) \\ &= 1 - 0.625 \\ &= 0.375 \end{aligned}$$

# Contoh

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

Outlook	Decision		
	Yes	No	Total
Overcast	4	0	4
Rain	4	1	5
Sunny	2	3	5
Total	7	7	14
Temp.	Decision		
	Yes	No	Total
Cool	0	4	4
Hot	2	2	4
Mild	2	4	6
Total	4	10	14

Humidity	Decision		
	Yes	No	Total
High	4	3	7
Normal	7	0	7
Total	11	3	14
Wind	Decision		
	Yes	No	Total
Weak	2	6	8
Strong	4	2	6
Total	6	8	14

## Hitung entropy outlook

$$E(\text{Decision} | \text{Outlook} = \text{Overcast}) = \left( -\frac{0}{4} \times \log_2 \left( \frac{0}{4} \right) \right) + \left( -\frac{4}{4} \times \log_2 \left( \frac{4}{4} \right) \right) = 0$$

$$E(\text{Decision} | \text{Outlook} = \text{Rain}) = \left( -\frac{1}{5} \times \log_2 \left( \frac{1}{5} \right) \right) + \left( -\frac{4}{5} \times \log_2 \left( \frac{4}{5} \right) \right) \approx 0.722$$

$$E(\text{Decision} | \text{Outlook} = \text{Sunny}) = \left( -\frac{3}{5} \times \log_2 \left( \frac{3}{5} \right) \right) + \left( -\frac{2}{5} \times \log_2 \left( \frac{2}{5} \right) \right) \approx 0.971$$

$$E(\text{Decision} | \text{Outlook}) = \left( \frac{4}{14} \times 0 \right) + \left( \frac{5}{14} \times 0.722 \right) + \left( \frac{5}{14} \times 0.971 \right) \approx 0.605$$

## Hitung entropy temp

$$E(\text{Decision}|\text{Temp.} = \text{Cool}) = \left(-\frac{0}{4} \times \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} \times \log_2 \left(\frac{4}{4}\right)\right) = 0$$

$$E(\text{Decision}|\text{Temp.} = \text{Hot}) = \left(-\frac{2}{4} \times \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \times \log_2 \left(\frac{2}{4}\right)\right) = 1$$

$$E(\text{Decision}|\text{Temp.} = \text{Mid}) = \left(-\frac{2}{6} \times \log_2 \left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6} \times \log_2 \left(\frac{4}{6}\right)\right) \approx 0.918$$

$$E(\text{Decision}|\text{Temp.}) = \left(\frac{4}{14} \times 0\right) + \left(\frac{4}{14} \times 1\right) + \left(\frac{6}{14} \times 0.918\right) \approx 0.679$$

## Hitung entropy humidity

$$E(\text{Decision} | \text{Humidity} = \text{High}) \\ = \left( -\frac{4}{7} \times \log_2 \left( \frac{4}{7} \right) \right) + \left( -\frac{3}{7} \times \log_2 \left( \frac{3}{7} \right) \right) \approx 0.985$$

$$E(\text{Decision} | \text{Humidity} = \text{Normal}) \\ = \left( -\frac{0}{7} \times \log_2 \left( \frac{0}{7} \right) \right) + \left( -\frac{7}{7} \times \log_2 \left( \frac{7}{7} \right) \right) = 0$$

$$E(\text{Decision} | \text{Humidity}) = \left( \frac{7}{14} \times 0.985 \right) + \left( \frac{7}{14} \times 0 \right) = 0.493$$



## Hitung entropy wind

$$E(\text{Decision}|\text{Wind} = \text{Weak}) = \left(-\frac{2}{8} \times \log_2\left(\frac{2}{8}\right)\right) + \left(-\frac{6}{8} \times \log_2\left(\frac{6}{8}\right)\right)$$

$$E(\text{Decision}|\text{Wind} = \text{Strong}) = \left(-\frac{4}{6} \times \log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} \times \log_2\left(\frac{2}{6}\right)\right) \approx 0.918$$

$$E(\text{Decision}|\text{Wind}) = \left(\frac{8}{14} \times 0.811\right) + \left(\frac{6}{14} \times 0.918\right) \approx 0.857$$

## Hitung entropy decision

$$E(\textit{Decision}) = \left( -\frac{4}{14} \times \log_2 \left( \frac{4}{14} \right) \right) + \left( -\frac{10}{14} \times \log_2 \left( \frac{10}{14} \right) \right) \approx 0.863$$

# Hitung IG

$$IG(Y, X) = E(Y) - E(Y|X)$$

$$\begin{aligned}IG(Decision, Outlook) &= E(Decision) - E(Decision|Outlook) \\&= 0.863 - 0.605 \\&= 0.259\end{aligned}$$

$$\begin{aligned}IG(Decision, Temp.) &= E(Decision) - E(Decision|Temp.) \\&= 0.863 - 0.679 \\&= 0.184\end{aligned}$$

$$\begin{aligned}IG(Decision, Humidity) &= E(Decision) - E(Decision|Humidity) \\&= 0.863 - 0.493 \\&= 0.371\end{aligned}$$

$$\begin{aligned}IG(Decision, Wind) &= E(Decision) - E(Decision|Wind) \\&= 0.863 - 0.857 \\&= 0.006\end{aligned}$$

# Hasil

atribut	Decision			Entropy	Information Gain
	Yes	No	Total		
Outlook					
• Overcast	4	0	4	0	0.259
• Rain	4	1	5	0.722	
• Sunny	2	3	5	0.971	
Temp.					
• 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	0.811	0.006
• Strong	4	2	6	0.918	

