

DATA MINING

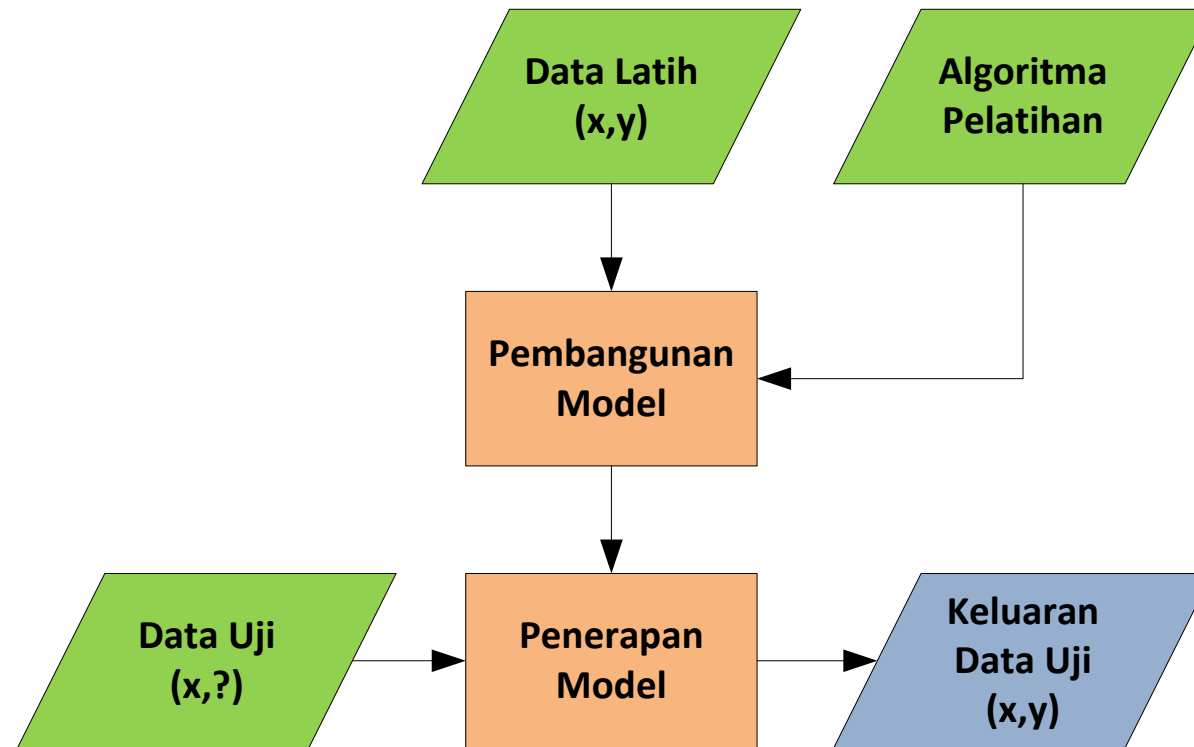
3 SKS | Semester 6 | S1 Sistem Informasi

TEKNIK KLASIFIKASI – ALGORITMA C4.5

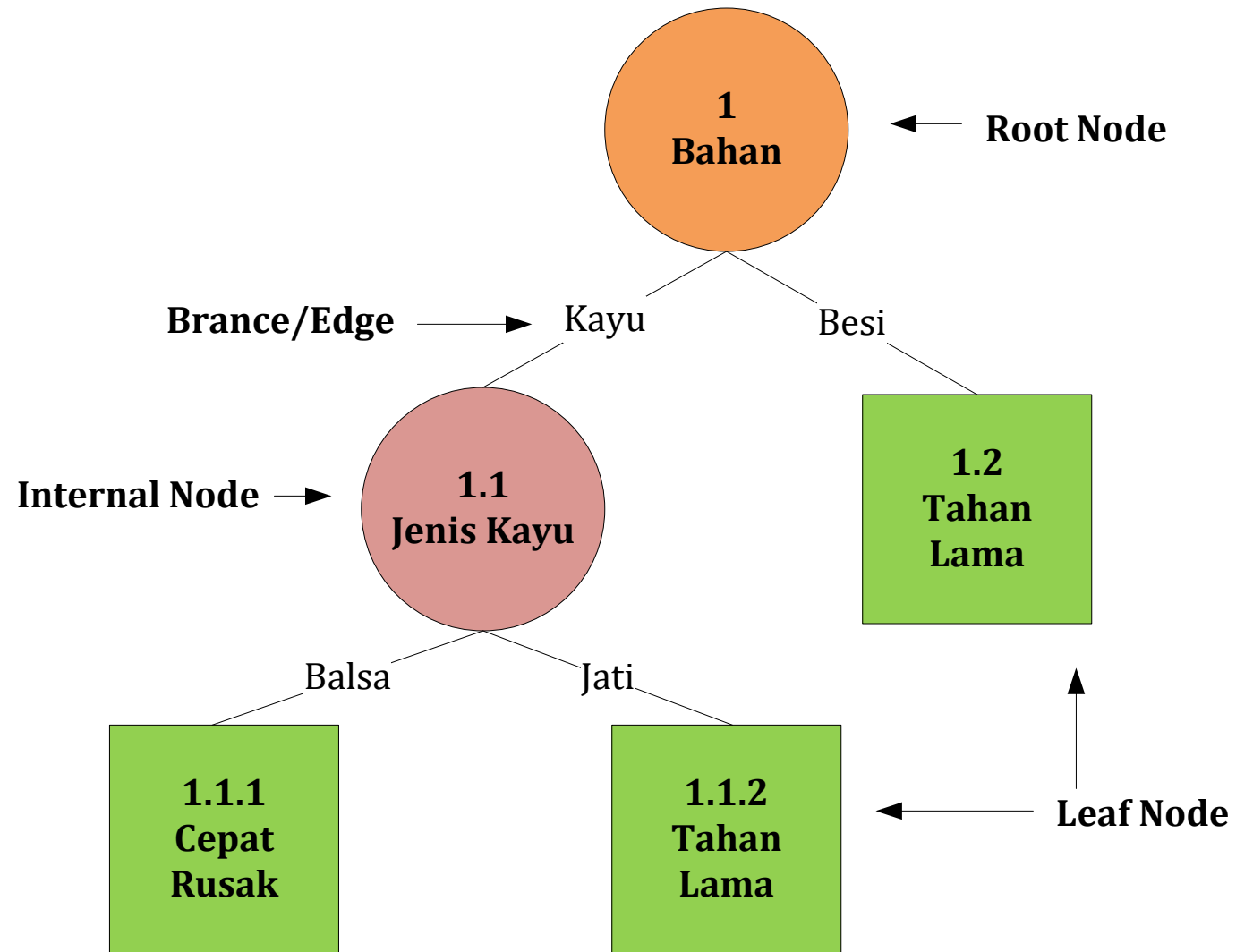
Nizar Rabbi Radliya
nizar@email.unikom.ac.id



TEKNIK KLASIFIKASI



DECISION TREE



ALGORITMA C4.5

Pembentukan decision tree menggunakan C4.5:

1. Pilih atribut sebagai akar;
2. Buat cabang untuk tiap-tiap nilai;
3. Bagi kasus dalam cabang;
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pemilihan atribut sebagai akar, didasarkan pada nilai **gain** tertinggi.

ALGORITMA C4.5

$$\mathbf{Gain(S,A)} = \mathbf{Entropy(S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \mathbf{Entropy(S_i)}$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke-i (S_i)

$|S|$: jumlah kasus dalam S

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

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

CONTOH KASUS**Data Latih: Keputusan Bermain Tennis**

No	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Cloudy	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Cool	Normal	True	Yes
7	Cloudy	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Cloudy	Mild	High	True	Yes
13	Cloudy	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

CONTOH KASUS

Perhitungan Node 1

Node			Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1	Total		14	4	10	0.863	
	Outlook						0.259
		Cloudy	4	0	4	0.000	
		Rainy	5	1	4	0.722	
		Sunny	5	3	2	0.971	
	Temperature						0.184
		Cool	4	0	4	0.000	
		Hot	4	2	2	1.000	
		Mild	6	2	4	0.918	
	Humidity						0.371
		High	7	4	3	0.985	
		Normal	7	0	7	0.000	
	Windy						0.006
		False	8	2	6	0.811	
		True	6	4	2	0.918	

CONTOH KASUS

Perhitungan Node 1

Node			Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1	Total		14	4	10	0.863	
	Outlook						0.259
		Cloudy	4	0	4	0.000	
		Rainy	5	1	4	0.722	
		Sunny	5	3	2	0.971	

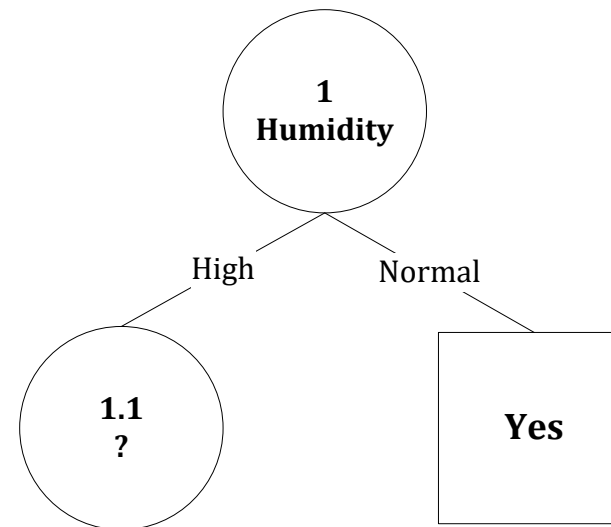
$$\begin{aligned}
 \text{Entropy}(S) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 \text{Entropy}(\text{Total}) &= \left(-\frac{S_1}{S} * \log_2\left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2\left(\frac{S_2}{S}\right)\right) \\
 \text{Entropy}(\text{Total}) &= \left(-\frac{4}{14} * \log_2\left(\frac{4}{14}\right)\right) + \left(-\frac{10}{14} * \log_2\left(\frac{10}{14}\right)\right) \\
 \text{Entropy}(\text{Total}) &= 0.863
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(S,A) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \\
 \text{Gain}(\text{Total}, \text{Outlook}) &= \text{Entropy}(\text{Total}) - \sum_{i=1}^n \frac{|\text{Outlook}_i|}{|\text{Total}|} * \text{Entropy}(\text{Outlook}_i) \\
 \text{Gain}(\text{Total}, \text{Outlook}) &= 0.863 - \left(\left(\frac{4}{14} * 0\right) + \left(\frac{5}{14} * 0.722\right) + \left(\frac{5}{14} * 0.971\right)\right) = 0.259
 \end{aligned}$$

CONTOH KASUS

Pohon Keputusan hasil Perhitungan Node 1

	Humidity						0.371
		High	7	4	3	0.985	
		Normal	7	0	7	0.000	



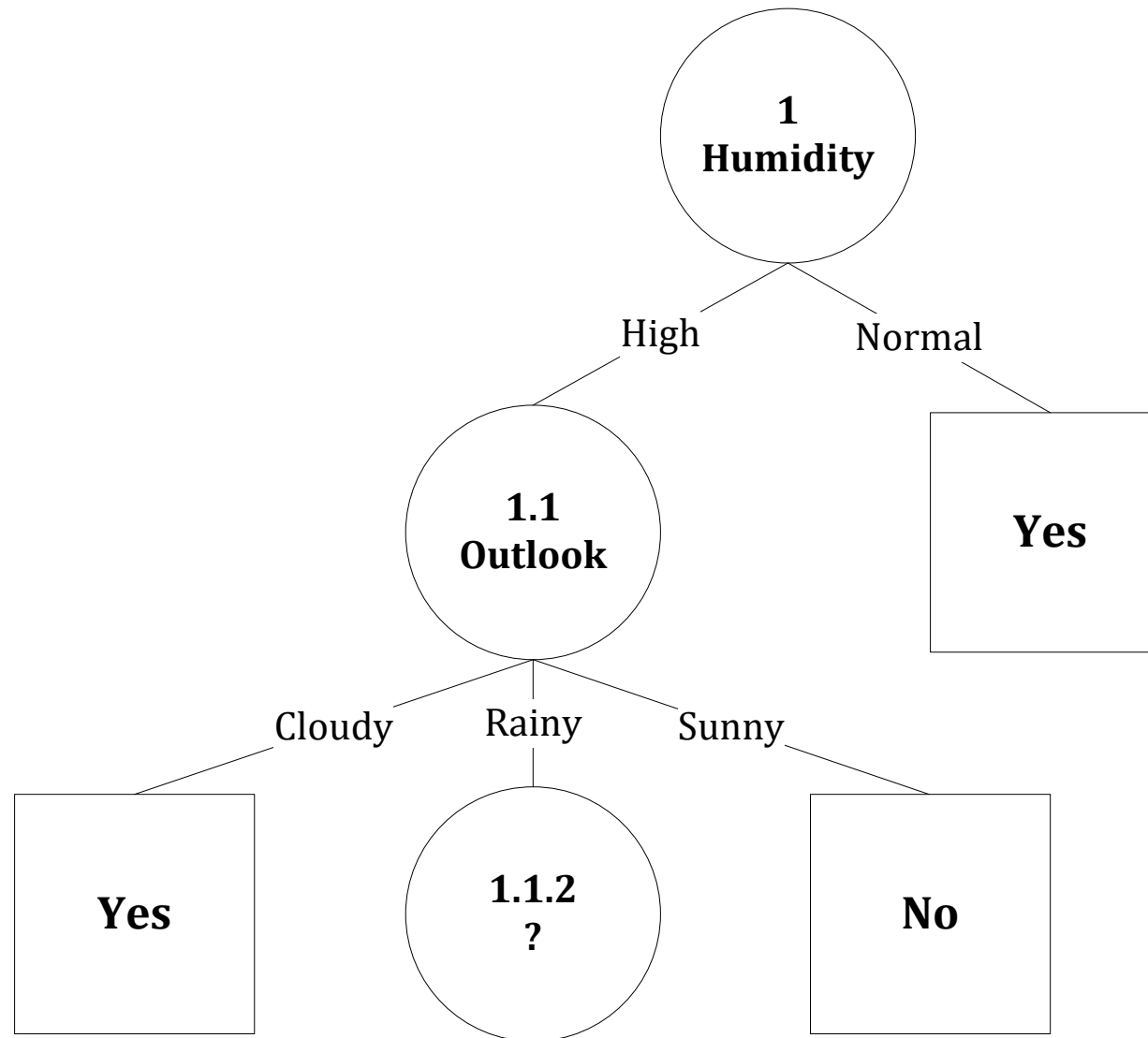
CONTOH KASUS

Perhitungan Node 1.1

Node			Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1.1	Humidity:High		7	4	3	0.985	
	Outlook						0.700
		Cloudy	2	0	2	0.000	
		Rainy	2	1	1	1.000	
		Sunny	3	3	0	0.000	
	Temperature						0.020
		Cool	0	0	0	0.000	
		Hot	3	2	1	0.918	
		Mild	4	2	2	1.000	
	Windy						0.020
		False	4	2	2	1.000	
		True	3	2	1	0.918	

CONTOH KASUS

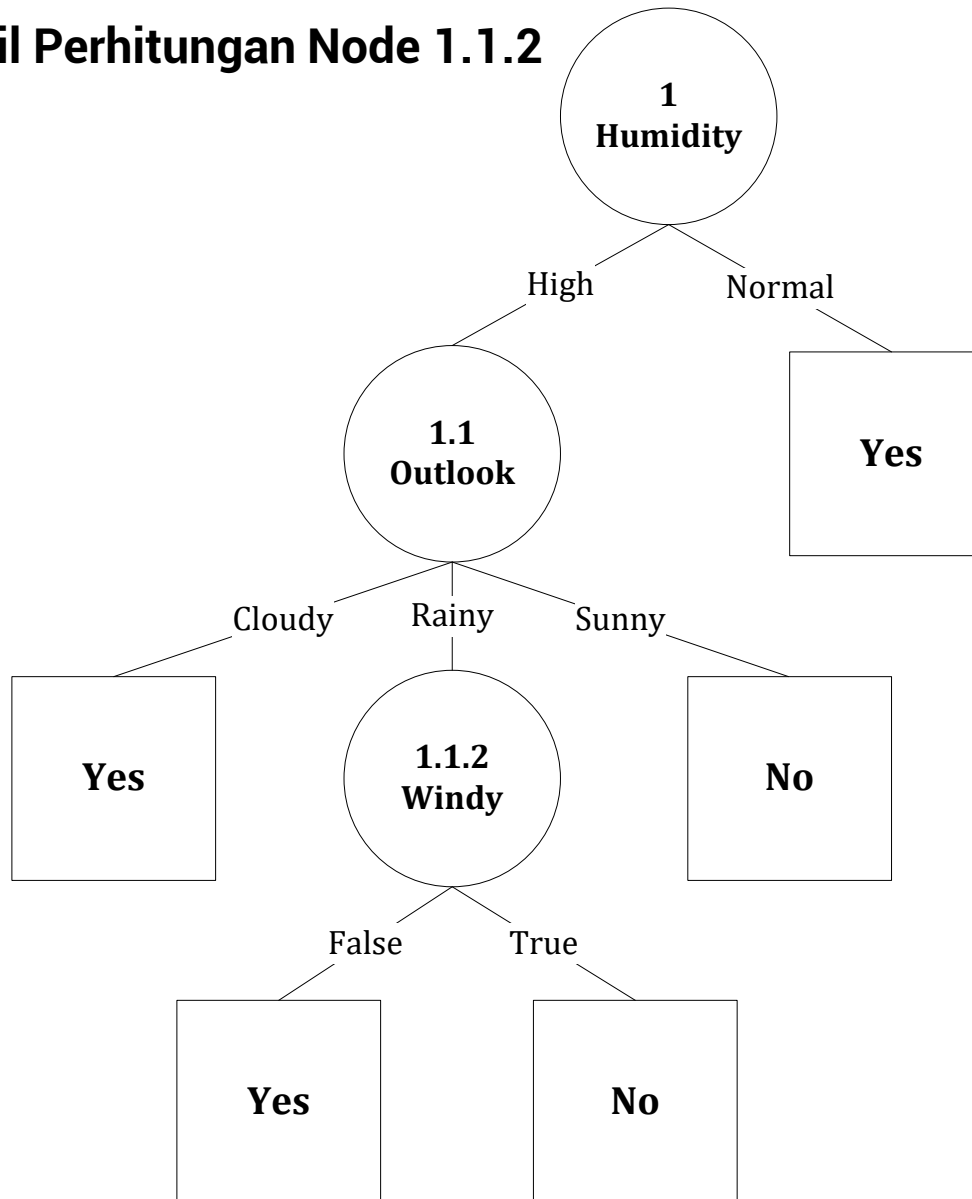
Pohon Keputusan hasil Perhitungan Node 1.1



CONTOH KASUS

Perhitungan Node 1.1.2

Node			Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1.1.2	Humidity:High & Outlook:Rainy		2	1	1	1.000	
	Temperature						0.000
		Cool	0	0	0	0.000	
		Hot	0	0	0	0.000	
		Mild	2	1	1	1.000	
	Windy						1.000
		False	1	0	1	0.000	
		True	1	1	0	0.000	

CONTOH KASUS**Pohon Keputusan hasil Perhitungan Node 1.1.2**

CONTOH KASUS**Mengubah model keputusan menjadi rule**

```
If (humidity = "normal") {  
    play = "yes";  
} else if (humidity = "high") {  
    if (outlook = "cloudy") {  
        play = "yes";  
    } else if (outlook = "sunny") {  
        play = "no";  
    } else if (outlook = "rainy") {  
        if (windy = "false") {  
            play = "yes";  
        } else if (windy = "true") {  
            play = "no";  
        }  
    }  
}  
}
```

LATIHAN

Buatlah pohon keputusan menggunakan teknik klasifikasi (dengan algoritma C4.5) untuk prediksi penerimaan dosen baru di program studi Sistem Informasi. Berikut di bawah ini adalah data latih yang akan digunakan.

No	Narkoba	Pendidikan S2	Toefl	Hasil
1	Negatif	S2 IF	≥ 500	Diterima
2	Positif	S2 SI	≥ 500	Diterima
3	Negatif	S2 DKV	< 500	Ditolak
4	Negatif	S2 IF	≥ 500	Diterima
5	Positif	S2 IF	≥ 500	Ditolak
6	Negatif	S2 IF	< 500	Ditolak
7	Positif	S2 SI	≥ 500	Diterima
8	Negatif	S2 SI	≥ 500	Diterima
9	Negatif	S2 DKV	≥ 500	Ditolak
10	Negatif	S2 IF	< 500	Ditolak

NEXT

TEKNIK KLASIFIKASI

