



TEKNIK INFORMATIKA
FAKULTAS TEKNIK UNIVERSITAS MATARAM

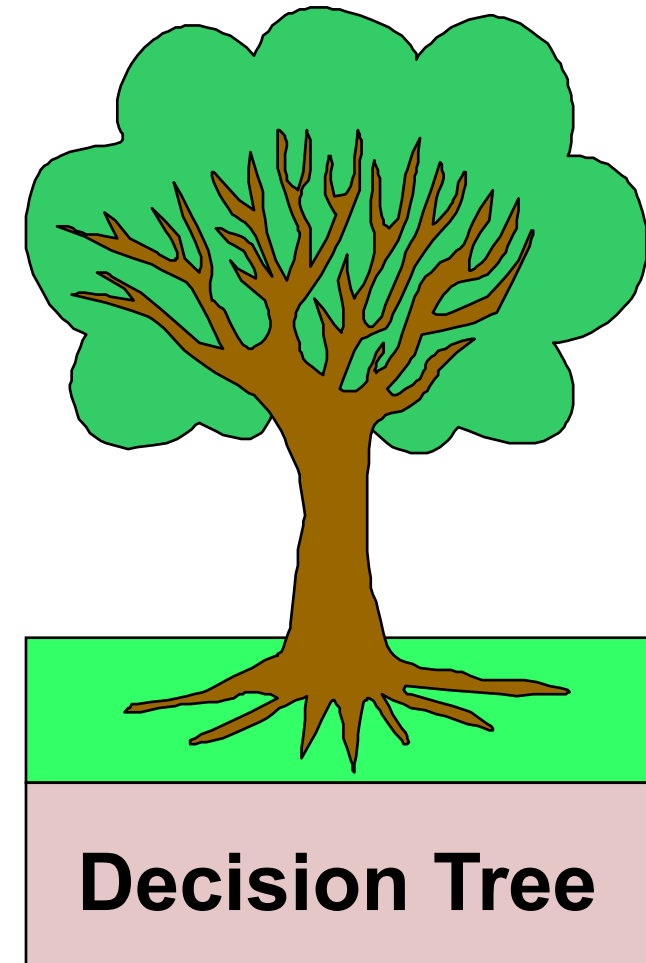


Decision Tree

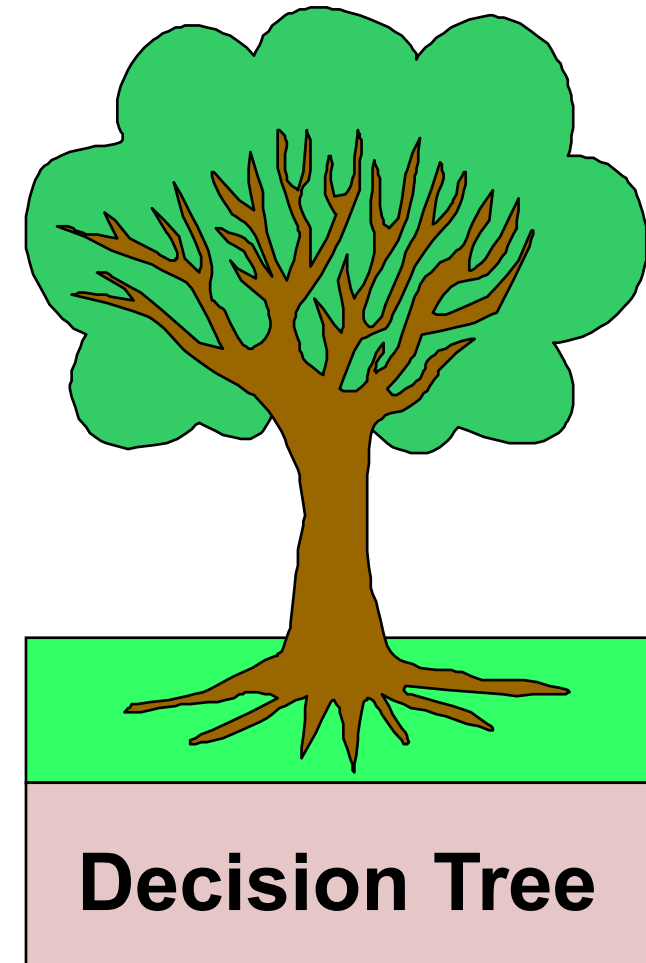
Ramaditia D

- Pendahuluan
- Contoh Decision Tree
- Proses Membentuk Pohon
- Metode ID3

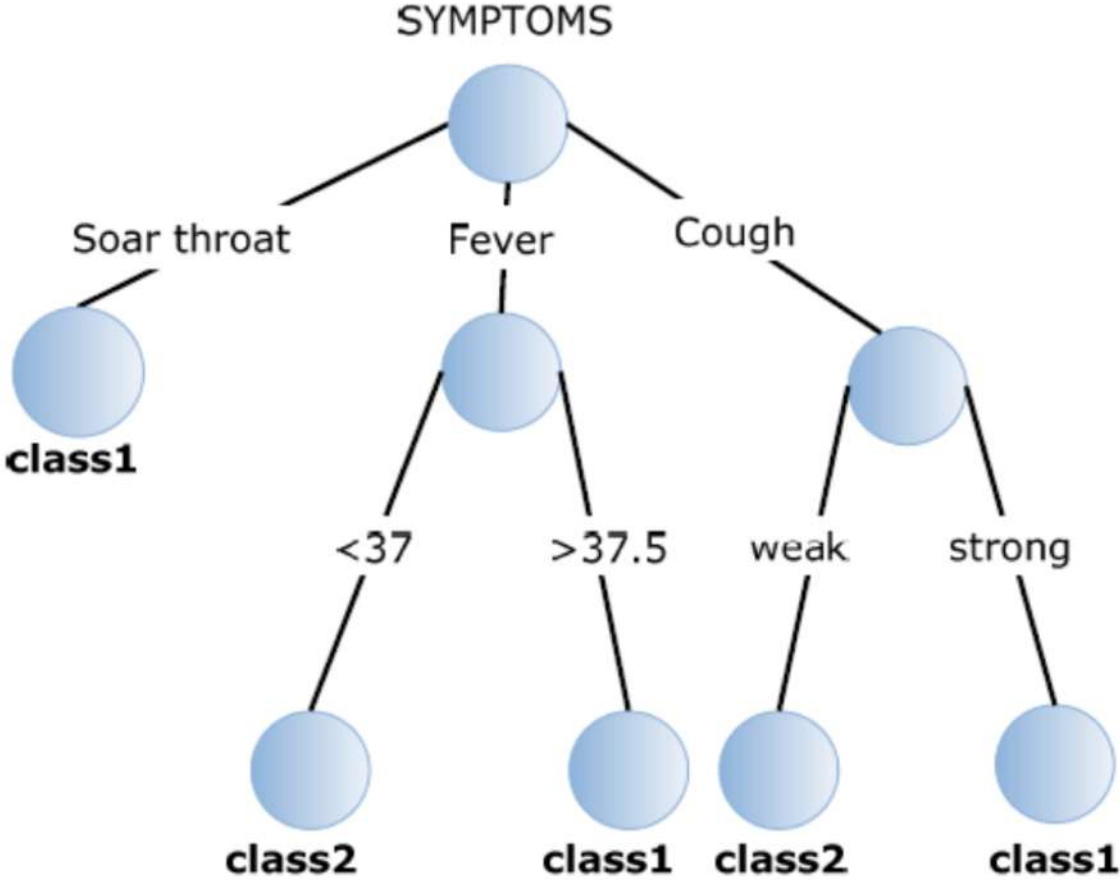
- Merupakan representasi visual berupa struktur pohon dari suatu pilihan aksi atau kondisi
- Merupakan mekanisme untuk menyederhanakan situasi kompleks ke dalam skenario yang lebih mudah dimengerti



- Pembelajaran Decision Tree (DT) adalah metode memperkirakan fungsi target bernilai diskret, dimana fungsi tersebut disajikan sebagai pohon keputusan.
- Representasi Pohon Keputusan:
 - Setiap Node internal menguji suatu atribut
 - Setiap cabang terkait dengan nilai atribut
 - Setiap Node daun menunjukkan suatu klasifikasi

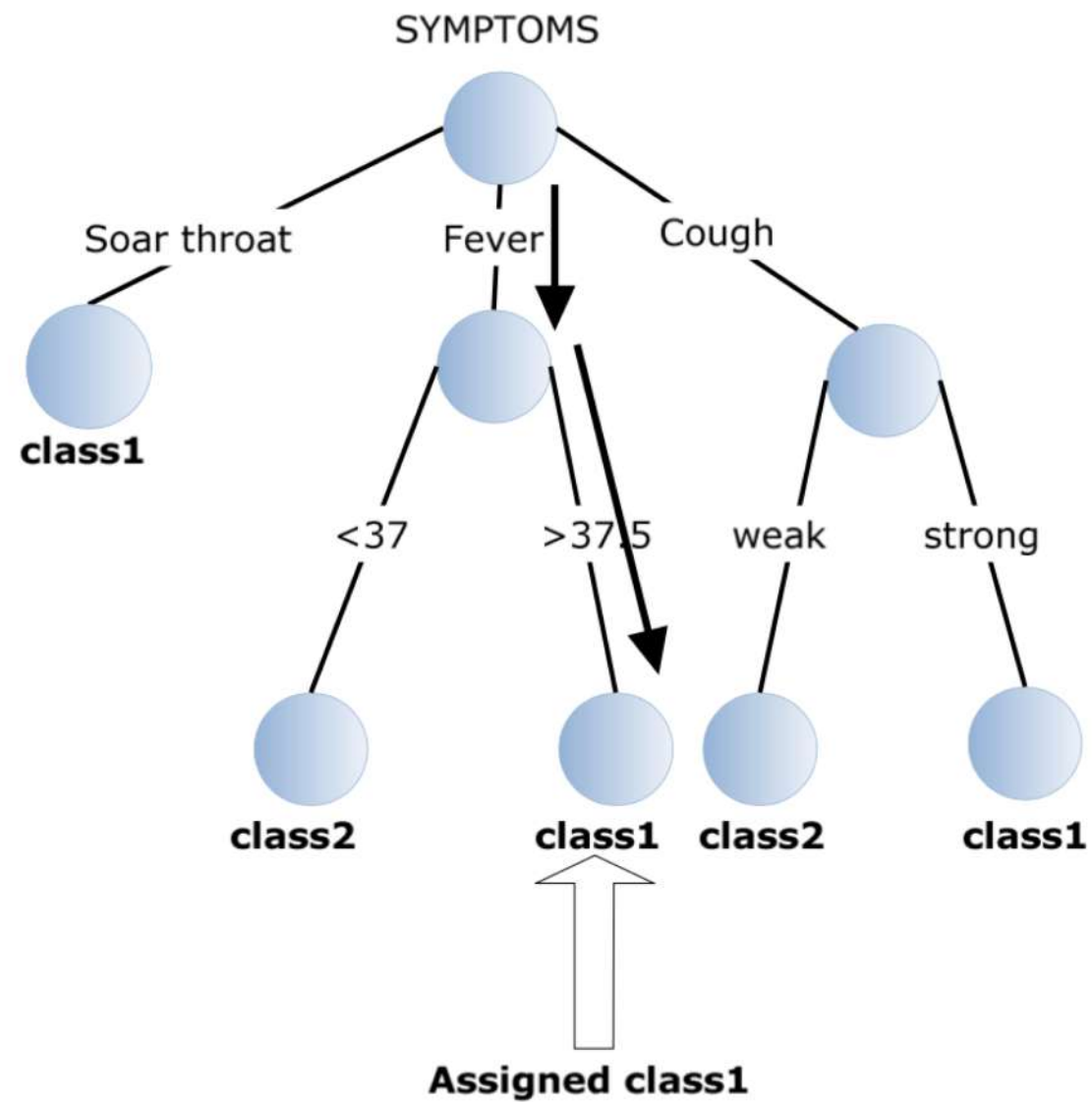


Flue (Class1) atau Tidak (Class2)?

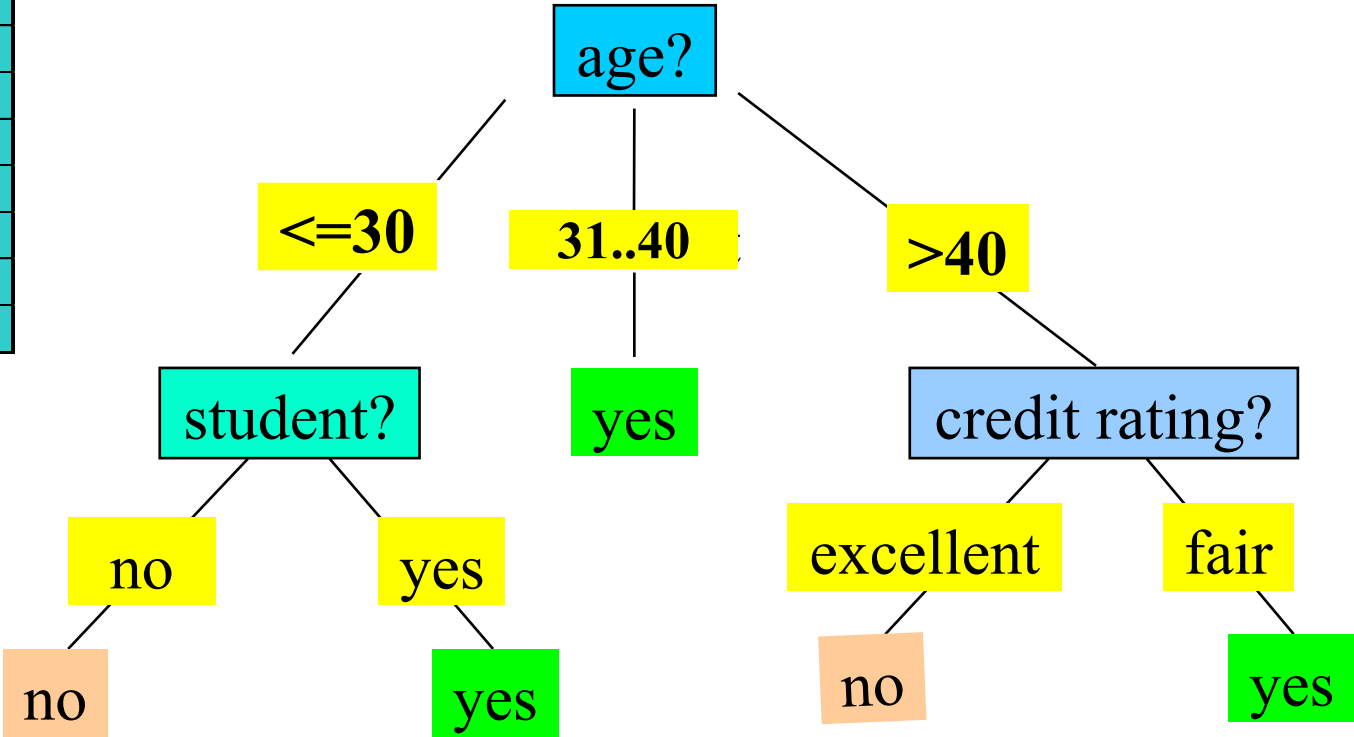


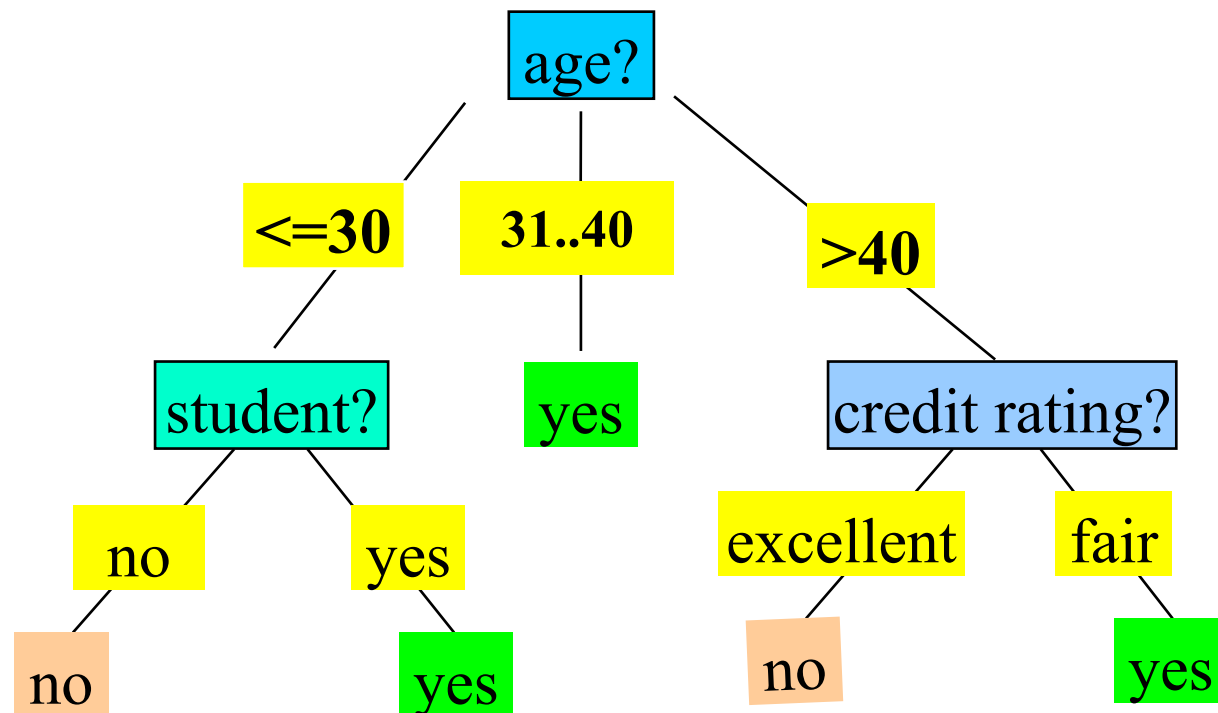
Flue (Class1) atau Tidak (Class2)?

Testing Data:
Pasien Demam (>37,5)



age	income	student	credit_rating	buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

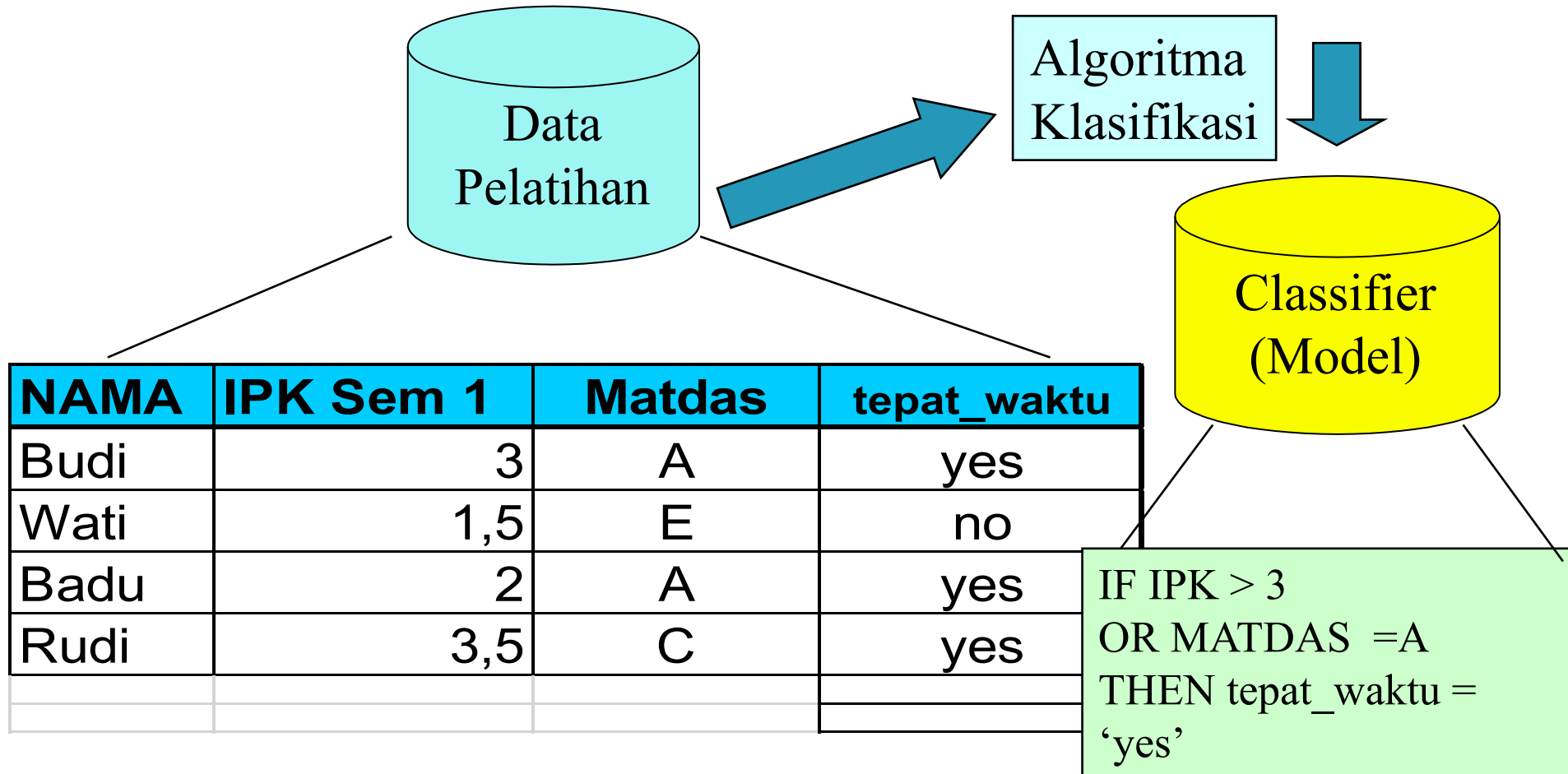


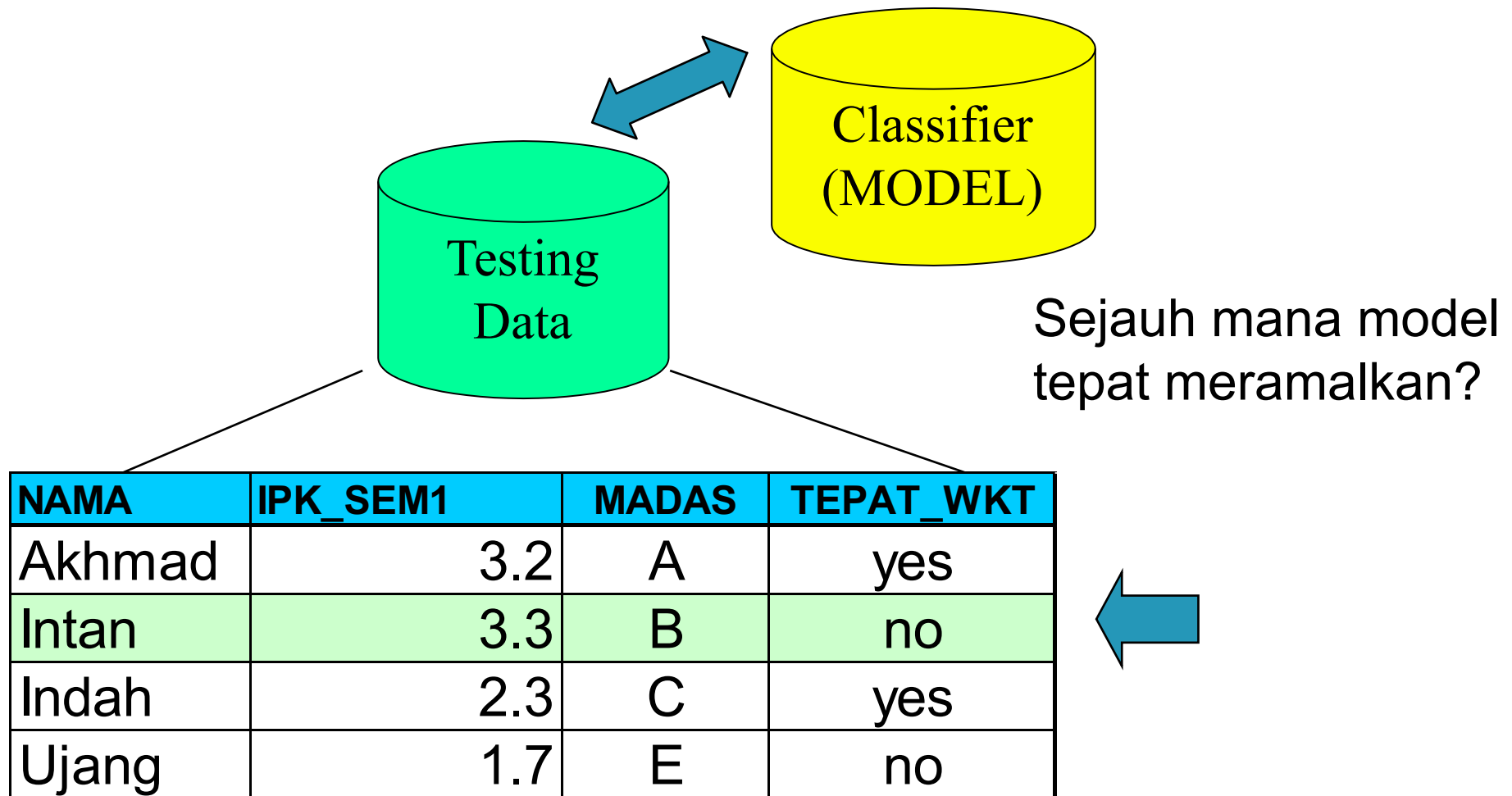


IF
 ((age<=30) and
 (student))
OR
 (age=31..40)
OR
 ((age>40) and
 (credit_rating=fair))
THEN
 BELI_PC=YES



- Input: data mahasiswa
 - Output: dua kelas (lulus_tepat_waktu **dan** lulus_terlambat)
- ✓ Jika diberikan data input mahasiswa, sistem secara otomatis menentukan mahasiswa tersebut akan lulus tepat waktu atau terlambat.

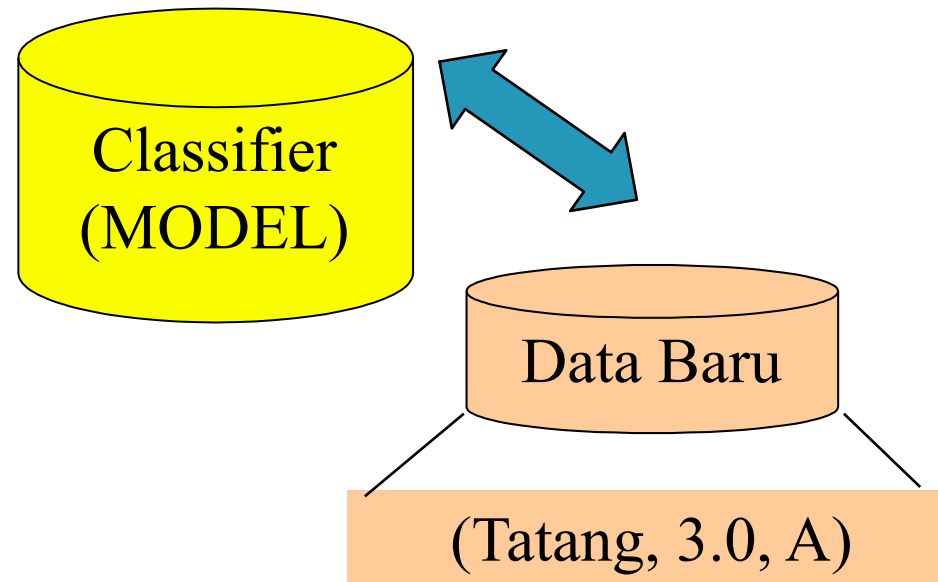






Proses Klasifikasi

11

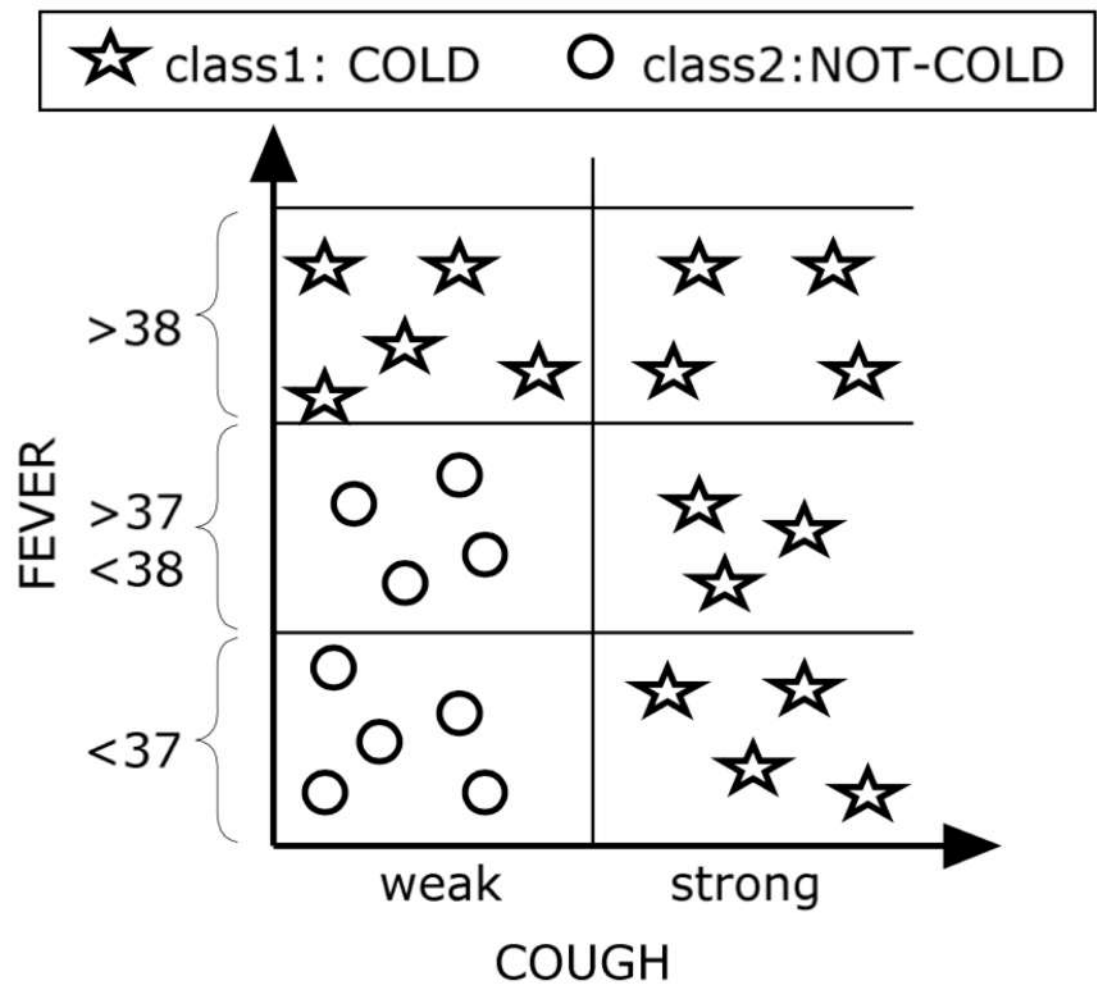


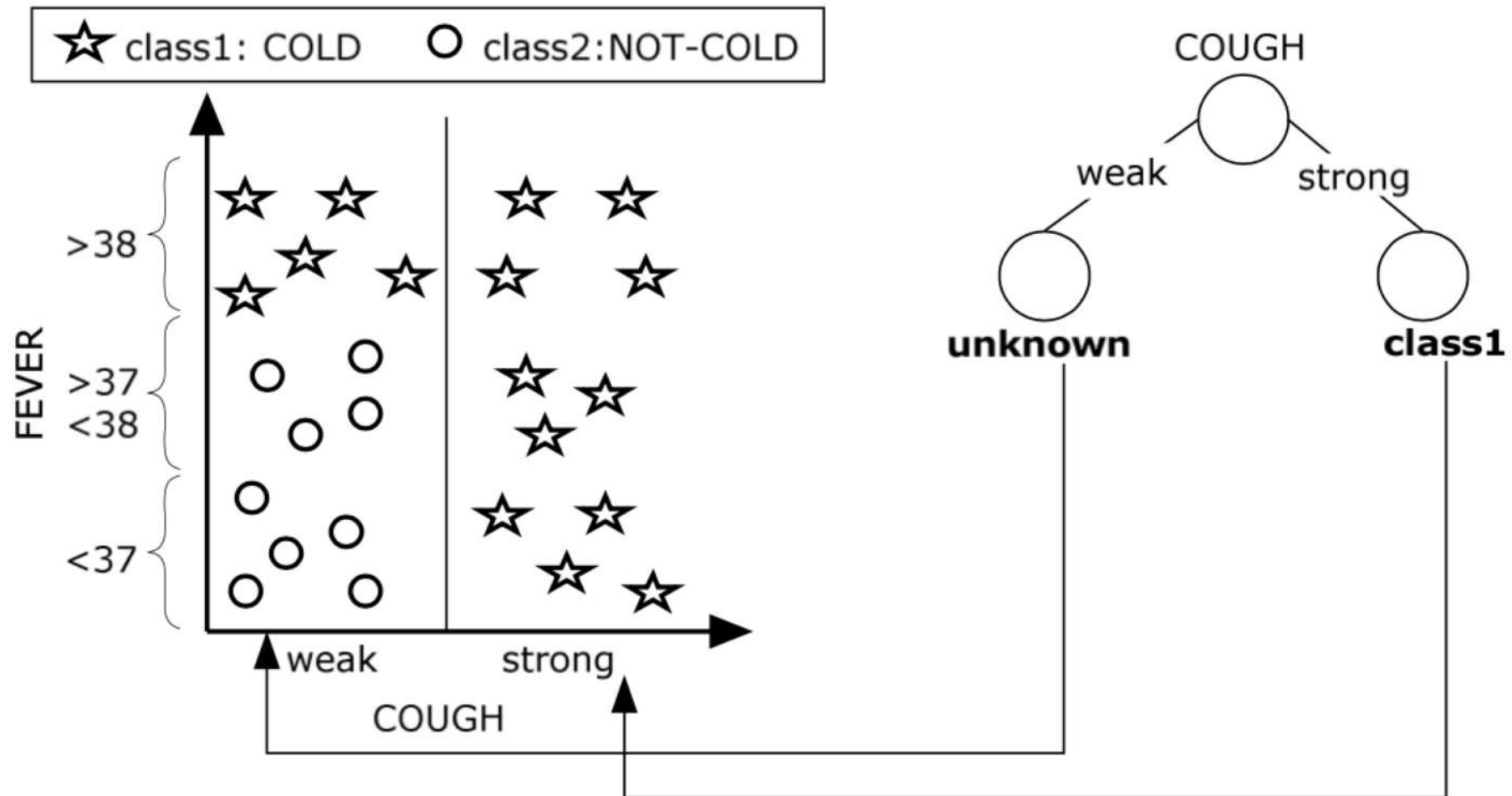
Lulus tepat waktu?

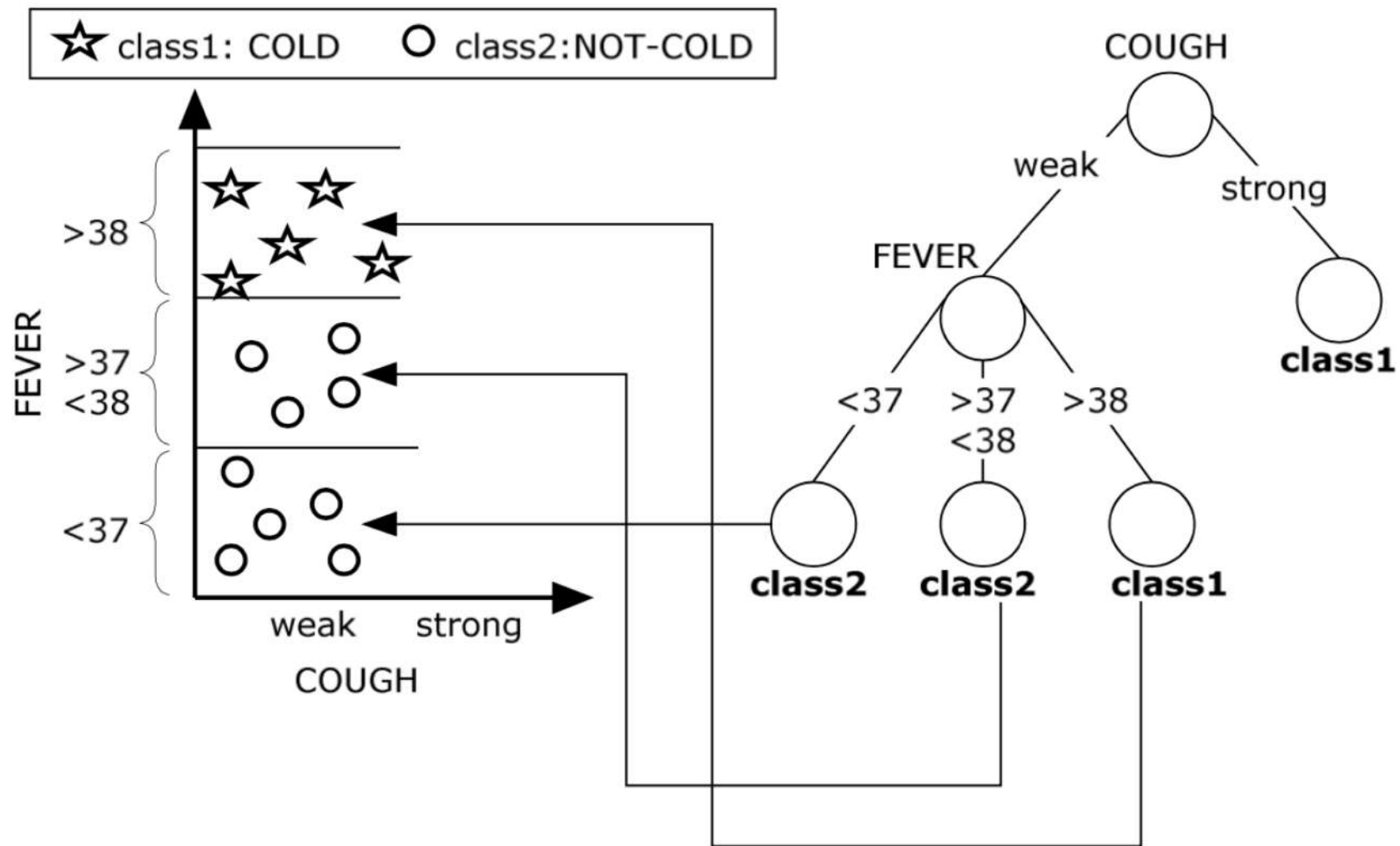


Yes

- Data Training (gambar):









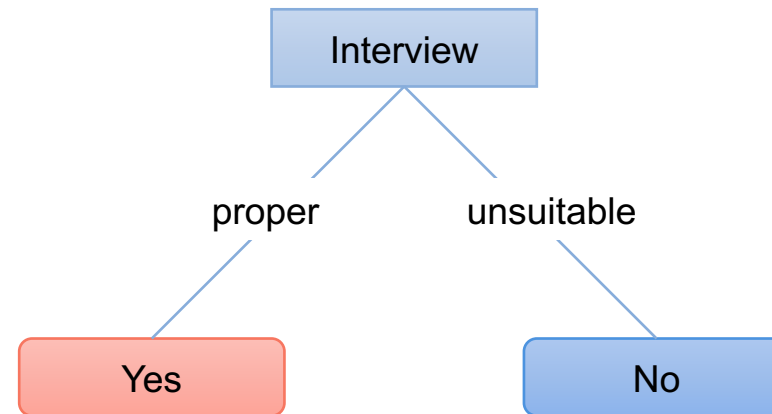
- Contoh: Data Rekrutmen Pegawai

Applicants	GPA	Psychology	Interview	Accepted
P1	Good	Strong	Proper	Yes
P2	Good	Moderate	Proper	Yes
P3	Good	Moderate	Unsuitable	No
P4	Good	Weak	Unsuitable	No
P5	Average	Strong	Proper	Yes
P6	Average	Moderate	Proper	Yes
P7	Average	Moderate	Unsuitable	No
P8	Average	Weak	Unsuitable	No
P9	Poor	Strong	Proper	Yes
P10	Poor	Moderate	Proper	Yes
P11	Poor	Moderate	Unsuitable	No
P12	Poor	Weak	Unsuitable	No



Bentuk Decision Tree

16



if Interview='Proper' then Accepted='Yes'



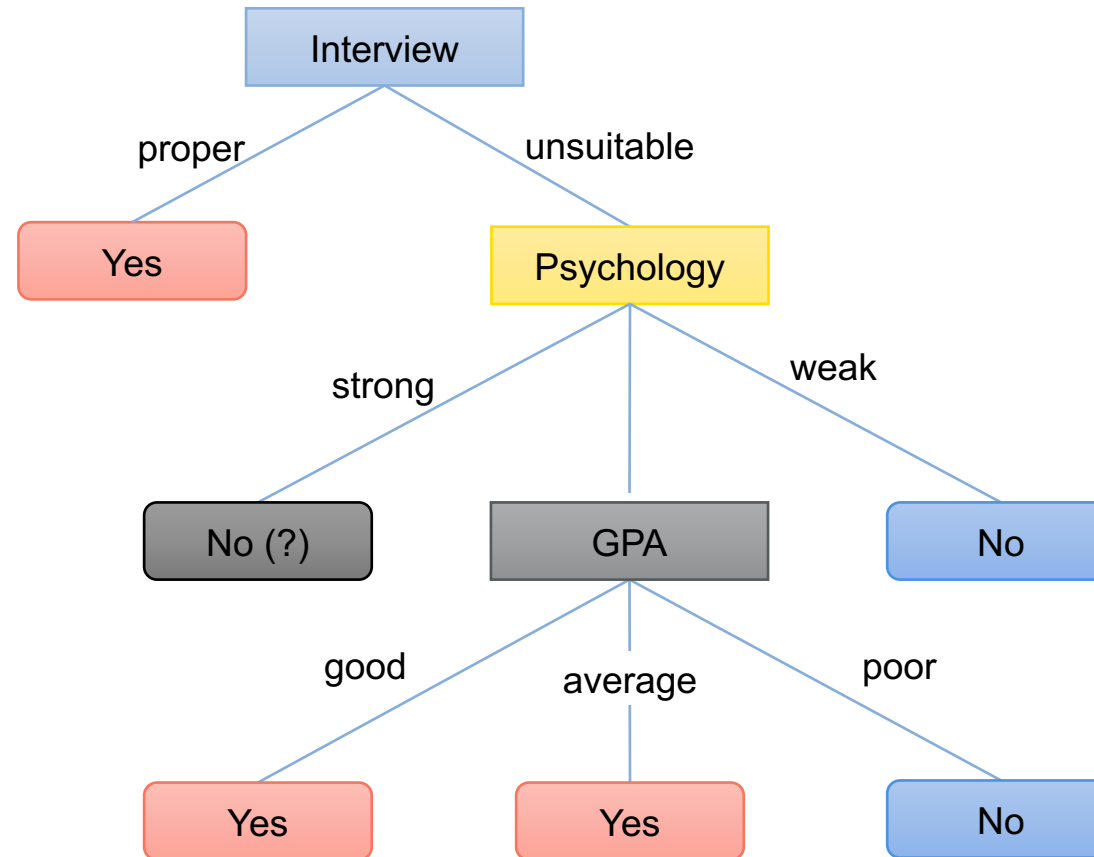
- Data Rekrutmen yang lain

Applicants	GPA	Psychology	Interview	Accepted
P1	Good	Strong	Proper	Yes
P2	Good	Moderate	Proper	Yes
P3	Good	Moderate	Unsuitable	Yes
P4	Good	Weak	Unsuitable	No
P5	Average	Strong	Proper	Yes
P6	Average	Moderate	Proper	Yes
P7	Average	Moderate	Unsuitable	Yes
P8	Average	Weak	Unsuitable	No
P9	Poor	Strong	Proper	Yes
P10	Poor	Moderate	Unsuitable	No
P11	Poor	Weak	Proper	Yes



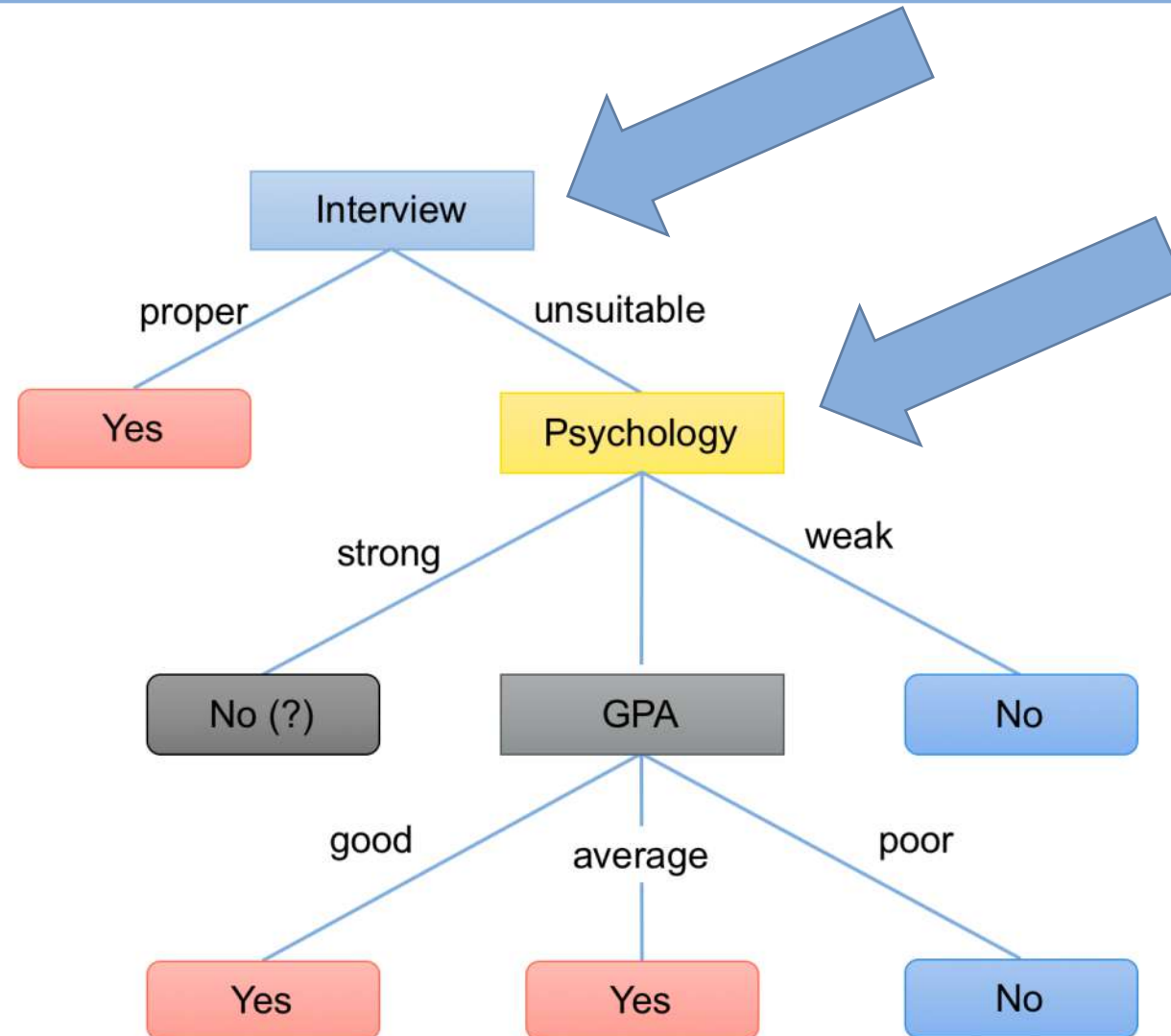
Bentuk Decision Tree

18



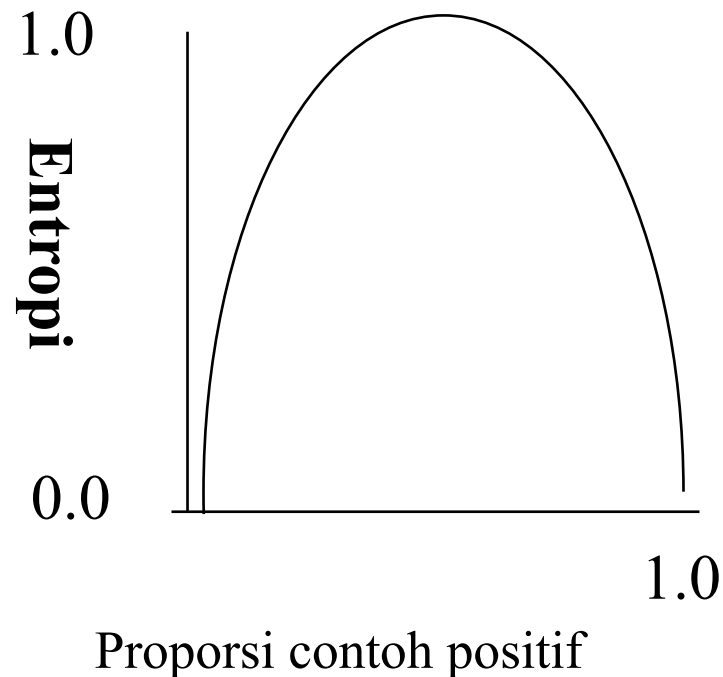
```
if (Interview='Proper') or  
  ( (Interview='Unsuitable')and(Psychology='Moderate')and(GPA='Good') ) or  
  ( (Interview='Unsuitable')and(Psychology='Moderate')and(GPA='Average') )  
then Accepted='Yes'
```

- A adalah atribut keputusan “terbaik” bagi node berikutnya
- Jadikan A sebagai atribut keputusan
- Untuk setiap nilai A, buat keturunan baru
- Urutkan sample pada node-node daun
- Jika sample telah terklasifikasi secara sempurna, STOP.
- Jika tidak, lakukan iterasi terhadap node daun baru.



- **Entropi**: Ukuran kemurnian, semakin murni, semakin homogen, semakin rendah nilainya.
- **Information Gain (IG)**: pengurangan entropi disebabkan oleh partisi berdasarkan suatu atribut.
- Semakin besar IG → atribut itu semakin membuat homogen → semakin bagus
- Ide Utama → pilih atribut dengan info gain yang paling besar

$$\text{Entropy}(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$



$$\begin{aligned} \text{Entropy}([9+,5-] \text{ ((9 positif, 5 neg))}) &= \\ &= -(9/14) \log_2(9/14) - (5/14) \log_2(5/14) \\ &= 0.940 \end{aligned}$$

$$\text{Entropy}([9+,5-]) = 0.940$$

$$\text{Entropy}([7+,7-]) = 1$$

$$\text{Entropy}([14+,0]) = 0$$

$$\text{Entroy}([0+,14-]) = 0$$

- $\text{Entropi}(S) = 0$, jika semua contoh pada S berada dalam kelas yang sama.
- $\text{Entropi}(S) = 1$, jika jumlah contoh positif dan jumlah contoh negatif dalam S adalah sama.
- $0 < \text{Entropi}(S) < 1$, jika jumlah contoh positif dan negatif dalam S tidak sama.

- Contoh Data:

Data	Class(Y/N)
1	Y
2	Y
3	Y
4	Y

$$\begin{aligned} Entropy(S) &= -p_y \log_2 p_y - p_n \log_2 p_n \\ &= -\frac{4}{4} \log_2 \left(\frac{4}{4}\right) - \frac{0}{4} \log_2 \left(\frac{0}{4}\right) \\ &= 0 \end{aligned}$$

Data	Class(Y/N)
1	Y
2	N
3	Y
4	N

$$\begin{aligned} Entropy(S) &= -p_y \log_2 p_y - p_n \log_2 p_n \\ &= -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right) \\ &= 1 \end{aligned}$$

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

Info (D) = Entrophy suatu data ‘D’ untuk kelas > 2



$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times I(D_j)$$

$$Gain(A) = Info(D) - Info_A(D)$$

Gain(A) menyatakan seberapa besar entropi berkurang akibat atribut A. Semakin besar semakin baik.

- Class P: buys_computer = “yes”
- Class N: buys_computer = “no”

$$Info(D) = I(9,5) = -\frac{9}{14} \log_2(\frac{9}{14}) - \frac{5}{14} \log_2(\frac{5}{14}) = 0.940$$

age	p _i	n _i	I(p _i , n _i)
<=30	2	3	0.971
31...40	4	0	0
>40	3	2	0.971

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

$$Info_{age}(D) = \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) = 0.694$$

$\frac{5}{14} I(2,3)$ berarti ada 5 dari 14
 “age <=30” dengan 2 yes
 dan 3 no.

$$Gain(Age) = Info(D) - Info_{age}(D) = 0.940 - 0.694 = 0.246$$

$$Gain(income) = 0.029$$

$$Gain(student) = 0.151$$

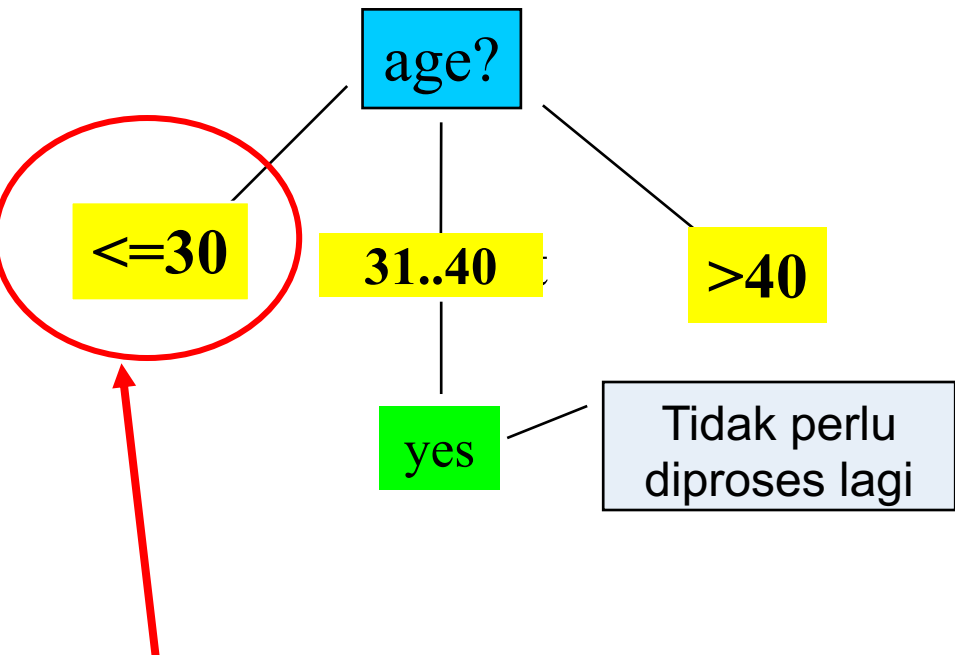
$$Gain(credit_rating) = 0.048$$

Gain (Age) = 0.246 ← yang terbesar, dipilih

Gain (income)=0.029

Gain(student)=0.151

Gain(credit_rating) =0.048



Selanjutnya... proses data yang <=30

Setelah AGE, atribut apa selanjutnya?

Diproses untuk setiap cabang selama masih ada > 1 kelas

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no
31...40	high	no	fair	yes
31...40	low	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes

Selanjutnya... proses data $\text{age} \leq 30$

age	income	student	credit_rating	buys_computer
≤ 30	high	no	fair	no
≤ 30	high	no	excellent	no
≤ 30	medium	no	fair	no
≤ 30	low	yes	fair	yes
≤ 30	medium	yes	excellent	yes

$$\text{Info}(D) = I(2,3) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) = 0.97$$

Gain(age) tidak perlu dihitung lagi,
hitung gain(student), gain(credit_rating), gain(income)

$$\text{Info}_{\text{student}}(D) = \frac{3}{5} I(0,3) + \frac{2}{5} I(2,0) = 0$$

$$\begin{aligned} \text{Gain}(\text{student}) &= \text{Info}(D) - \text{Info}_{\text{student}}(D) \\ &= 0.97 - 0 = 0.97 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

hitung gain(credit_rating)

$$Info(D) = I(2,3) = -\frac{2}{5}\log_2(\frac{2}{5}) - \frac{3}{5}\log_2(\frac{3}{5}) = 0.97$$

$$Info_{credit_rating}(D) = \frac{3}{5}I(1,2) + \frac{2}{5}I(1,1) = 0.95$$

Gain (credit_rating)

= Info(D) – Info_{credit_rating}(D)
= 0.97 – 0.95 = 0.02

$$Info_{income}(D) = \frac{2}{5}I(0,2) + \frac{2}{5}I(1,1) + \frac{1}{5}I(1,0) = 0.4$$

Gain (income)

= Info(D) – Info_{income}(D)
= 0.97 – 0.4 = 0.37



Bandingkan semua gain, ambil yang paling besar

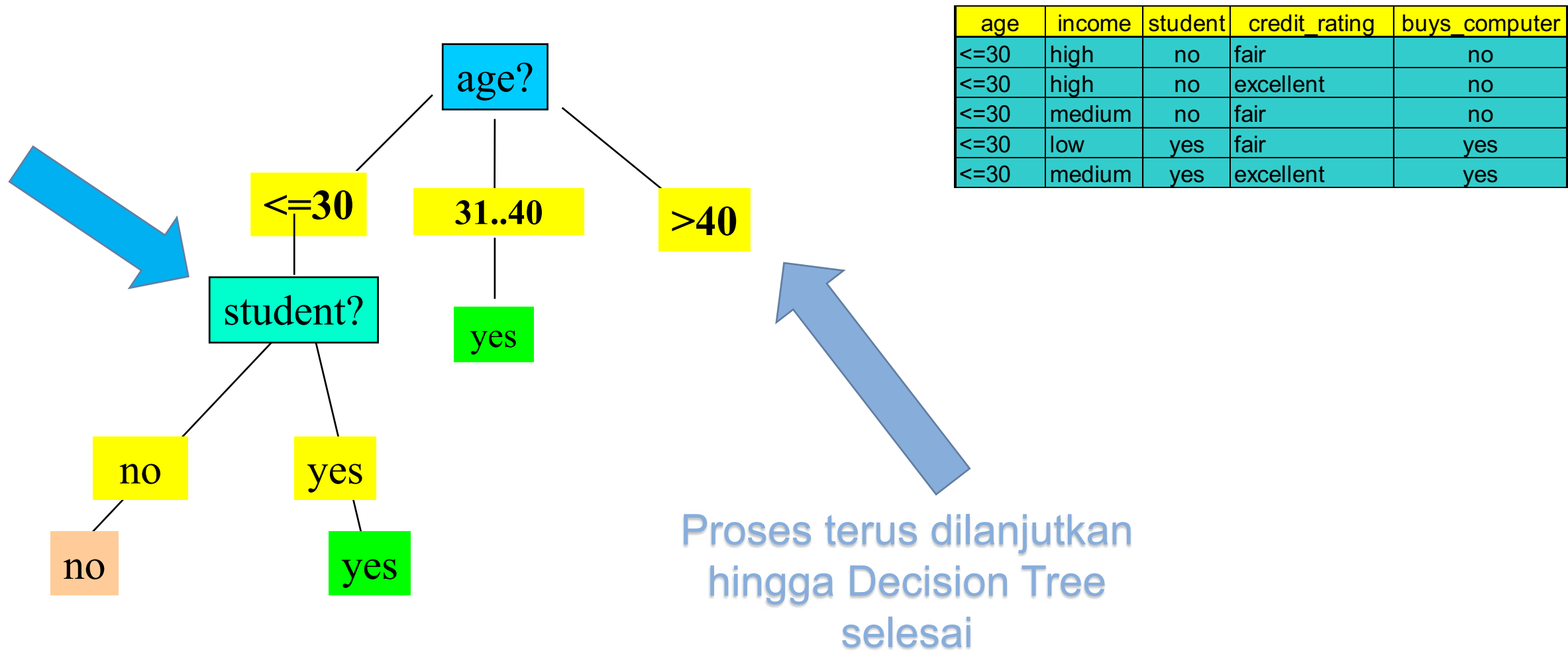
Gain (student) = 0.97

Gain (credit_rating) = 0.02

Gain (income) = 0.37



Paling besar:
student



Contoh: Data Rekrutmen Pegawai

Applicants	GPA	Psychology	Interview	Accepted
P1	Good	Strong	Proper	Yes
P2	Good	Moderate	Proper	Yes
P3	Good	Moderate	Unsuitable	Yes
P4	Good	Weak	Unsuitable	No
P5	Average	Strong	Proper	Yes
P6	Average	Moderate	Proper	Yes
P7	Average	Moderate	Unsuitable	Yes
P8	Average	Weak	Unsuitable	No
P9	Poor	Strong	Proper	Yes
P10	Poor	Moderate	Unsuitable	No
P11	Poor	Weak	Proper	Yes

Accepted
Yes
Yes
Yes
No
Yes
Yes
Yes
No
Yes
No
Yes

- $Entropy(A) = -p_y \log_2 p_y - p_n \log_2 p_n$
- $|A| = 11, \quad |A_y| = 8, \quad |A_n| = 3$
- $p_y = \frac{8}{11}, \quad p_n = \frac{3}{11}$
- $Entropy(A) = -\frac{8}{11} \log_2 \left(\frac{8}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right)$
- $= 0.8454$

GPA	Accepted
Good	Yes
Good	Yes
Good	Yes
Good	No
Average	Yes
Average	Yes
Average	Yes
Average	No
Poor	Yes
Poor	No
Poor	Yes

$$Gain(Accepted, GPA)$$

$$= Ent(A) - \frac{|A_g|}{|A|} Ent(A_g) - \frac{|A_a|}{|A|} Ent(A_a) - \frac{|A_p|}{|A|} Ent(A_p)$$

$|A_g| = [3y, 1n] = 4,$
 $|A_a| = [3y, 1n] = 4,$
 $|A_p| = [2y, 1n] = 3$

$Ent(A_g) = -p_{g,y} \log_2 p_{g,y} - p_{g,n} \log_2 p_{g,n}$
 $Ent(A_a) = -p_{a,y} \log_2 p_{a,y} - p_{a,n} \log_2 p_{a,n}$
 $Ent(A_p) = -p_{p,y} \log_2 p_{p,y} - p_{p,n} \log_2 p_{p,n}$

GPA	Accepted
Good	Yes
Good	Yes
Good	Yes
Good	No
Average	Yes
Average	Yes
Average	Yes
Average	No
Poor	Yes
Poor	No
Poor	Yes

$$|A_g| = [3y, 1n] = 4, |A_a| = [3y, 1n] = 4, |A_p| = [2y, 1n] = 3$$

$$\begin{aligned} Ent(A_g) &= -p_{g,y} \log_2 p_{g,y} - p_{g,n} \log_2 p_{g,n} \\ &= -\frac{3}{4} \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) = 0.8113 \end{aligned}$$

$$\begin{aligned} Ent(A_a) &= -p_{a,y} \log_2 p_{a,y} - p_{a,n} \log_2 p_{a,n} \\ &= -\frac{3}{4} \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) = 0.8113 \end{aligned}$$

$$\begin{aligned} Ent(A_p) &= -p_{p,y} \log_2 p_{p,y} - p_{p,n} \log_2 p_{p,n} \\ &= -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) = 0.9183 \end{aligned}$$

GPA	Accepted
Good	Yes
Good	Yes
Good	Yes
Good	No
Average	Yes
Average	Yes
Average	Yes
Average	No
Poor	Yes
Poor	No
Poor	Yes

$$|A_g| = 4, |A_a| = 4, |A_p| = 3$$

$$Ent(A_g) = 0.8113$$

$$Ent(A_a) = 0.8113$$

$$Ent(A_p) = 0.9183$$

$$\begin{aligned}
 &Gain(Accepted, GPA) \\
 &= Ent(A) - \frac{|A_g|}{|A|} Ent(A_g) - \frac{|A_a|}{|A|} Ent(A_a) - \frac{|A_p|}{|A|} Ent(A_p) \\
 &= 0.8454 - \frac{4}{11} 0.8113 - \frac{4}{11} 0.8113 - \frac{3}{11} 0.9183 \\
 &= 0.0049
 \end{aligned}$$

Psy	Accepted
Strong	Yes
Mod	Yes
Mod	Yes
Weak	No
Strong	Yes
Mod	Yes
Mod	Yes
Weak	No
Strong	Yes
Mod	No
Weak	Yes

$$Gain(Accepted, Psychology)$$

$$= Ent(A) - \frac{|A_s|}{|A|} Ent(A_s) - \frac{|A_m|}{|A|} Ent(A_m) - \frac{|A_w|}{|A|} Ent(A_w)$$

$$|A_s| = [3y, 0n] = 3,$$

$$|A_m| = [4y, 1n] = 5,$$

$$|A_w| = [1y, 2n] = 3$$

$$Ent(A_s) = -\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right) = 0$$

$$Ent(A_m) = -\frac{4}{5}\log_2\left(\frac{4}{5}\right) - \frac{1}{5}\log_2\left(\frac{1}{5}\right) = 0.7219$$

$$Ent(A_w) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.9183$$

Psy	Accepted
Strong	Yes
Mod	Yes
Mod	Yes
Weak	No
Strong	Yes
Mod	Yes
Mod	Yes
Weak	No
Strong	Yes
Mod	No
Weak	Yes

$$|A_s| = 3, |A_m| = 5, |A_w| = 3$$

$$Ent(A_s) = 0$$

$$Ent(A_m) = 0.7219$$

$$Ent(A_w) = 0.9183$$

$$\begin{aligned}
 &Gain(Accepted, Psychology) \\
 &= Ent(A) - \frac{|A_s|}{|A|} Ent(A_s) - \frac{|A_m|}{|A|} Ent(A_m) - \frac{|A_w|}{|A|} Ent(A_w) \\
 &= 0.8454 - \frac{3}{11} 0 - \frac{5}{11} 0.7219 - \frac{3}{11} 0.9183 \\
 &= 0.2669
 \end{aligned}$$

Intv	Accepted
Proper	Yes
Proper	Yes
Unsuit	Yes
Unsuit	No
Proper	Yes
Proper	Yes
Unsuit	Yes
Unsuit	No
Proper	Yes
Unsuit	No
Proper	Yes

Gain(Accepted, Interview)

$$= Ent(A) - \frac{|A_p|}{|A|} Ent(A_p) - \frac{|A_u|}{|A|} Ent(A_u)$$

$$|A_p| = [6y, 0n] = 6,$$

$$|A_u| = [2y, 3n] = 5,$$

$$Ent(A_p) = -\frac{6}{6} \log_2 \left(\frac{6}{6}\right) - \frac{0}{6} \log_2 \left(\frac{0}{6}\right) = 0$$

$$Ent(A_u) = -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \log_2 \left(\frac{3}{5}\right) = 0.9710$$

Intv	Accepted
Proper	Yes
Proper	Yes
Unsuit	Yes
Unsuit	No
Proper	Yes
Proper	Yes
Unsuit	Yes
Unsuit	No
Proper	Yes
Unsuit	No
Proper	Yes

$$|A_p| = [6y, 0n] = 6, |A_u| = [2y, 3n] = 5,$$

$$Ent(A_p) = 0$$

$$Ent(A_u) = 0.9710$$

$$Gain(Accepted, Interview)$$

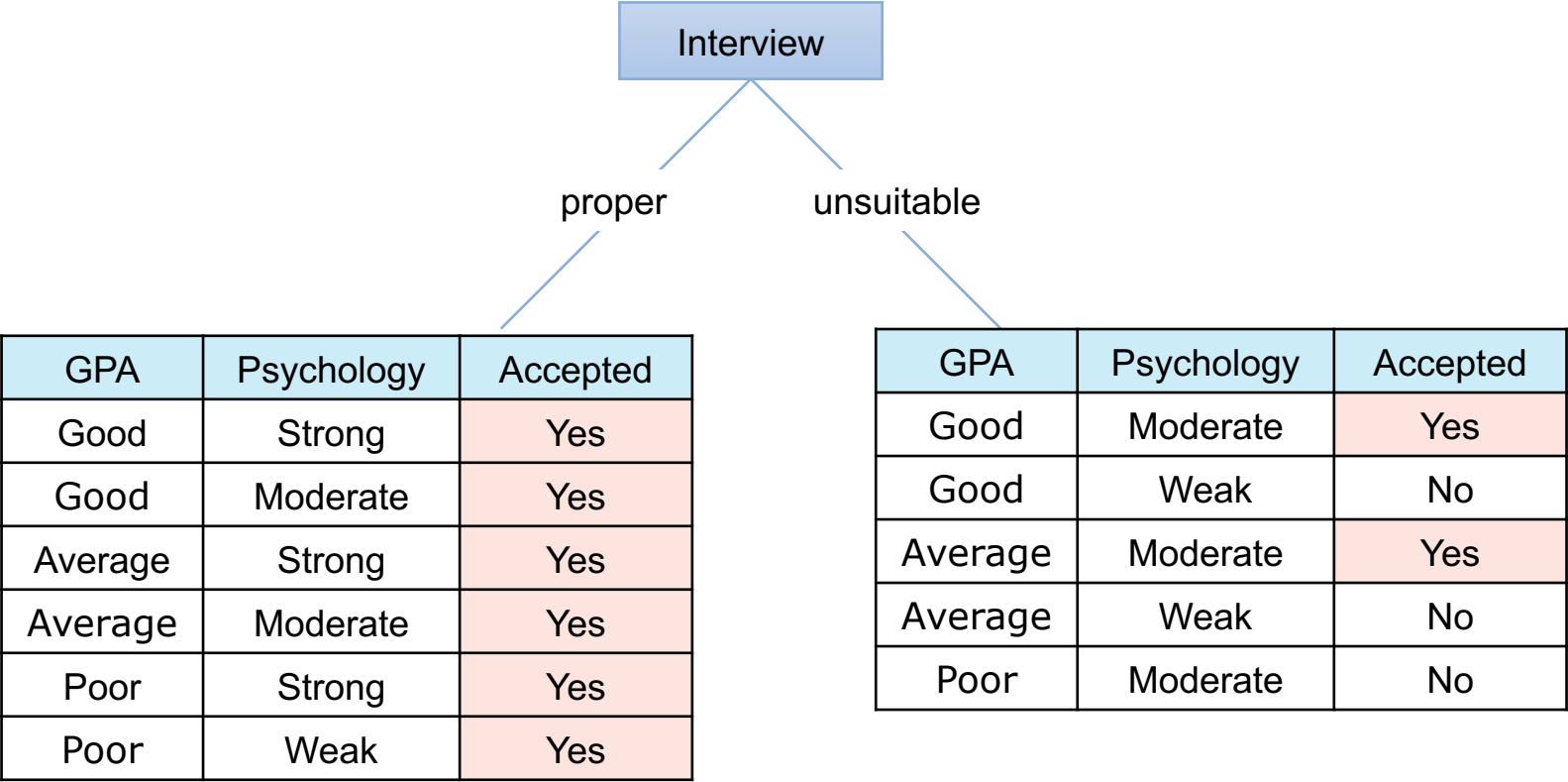
$$= Ent(A) - \frac{|A_p|}{|A|} Ent(A_p) - \frac{|A_u|}{|A|} Ent(A_u)$$

$$= 0.8454 - \frac{6}{11} 0 - \frac{5}{11} 0.9710$$

$$= 0.4040$$



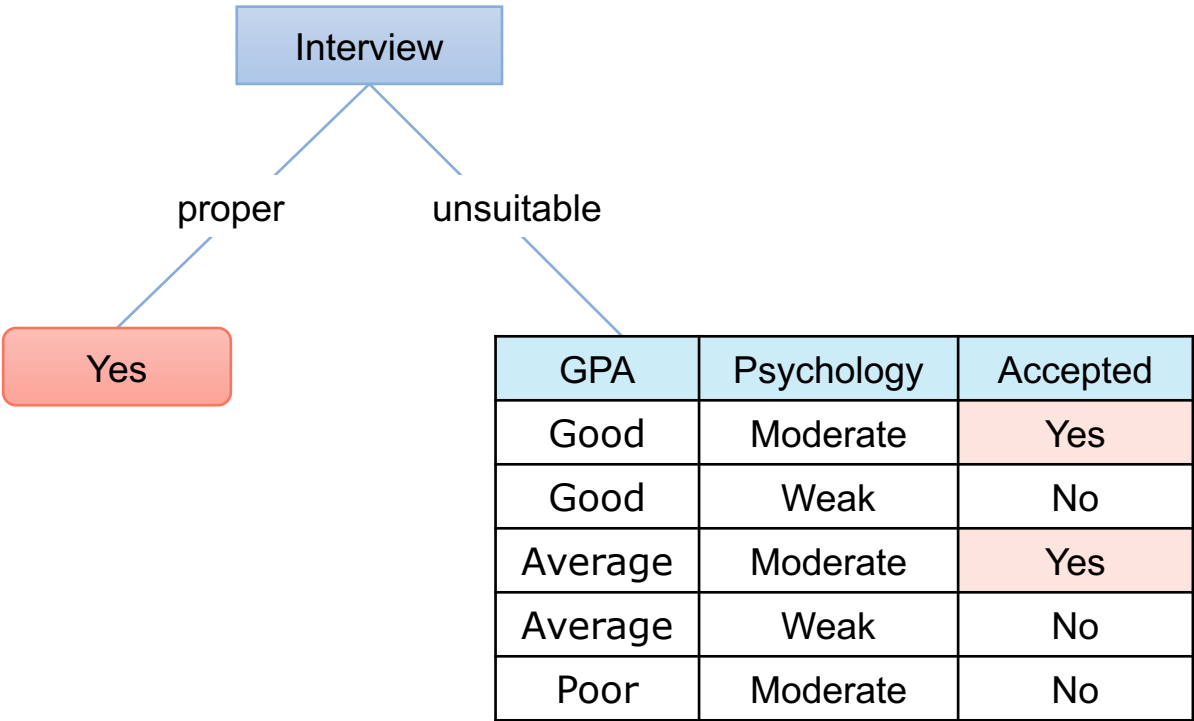
- $Gain(Accepted, GPA) = 0.0049$
- $Gain(Accepted, Psychology) = 0.2669$
- $Gain(Accepted, Interview) = 0.4040$
- Pilih Interview sebagai Root



Proper Interview

GPA	Psychology	Accepted
Good	Strong	Yes
Good	Moderate	Yes
Average	Strong	Yes
Average	Moderate	Yes
Poor	Strong	Yes
Poor	Weak	Yes

$Entropy(A) = 0$
Return node ("Yes")



GPA	Psychology	Accepted
Good	Moderate	Yes
Good	Weak	No
Average	Moderate	Yes
Average	Weak	No
Poor	Moderate	No

$$Ent(A) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.9709$$

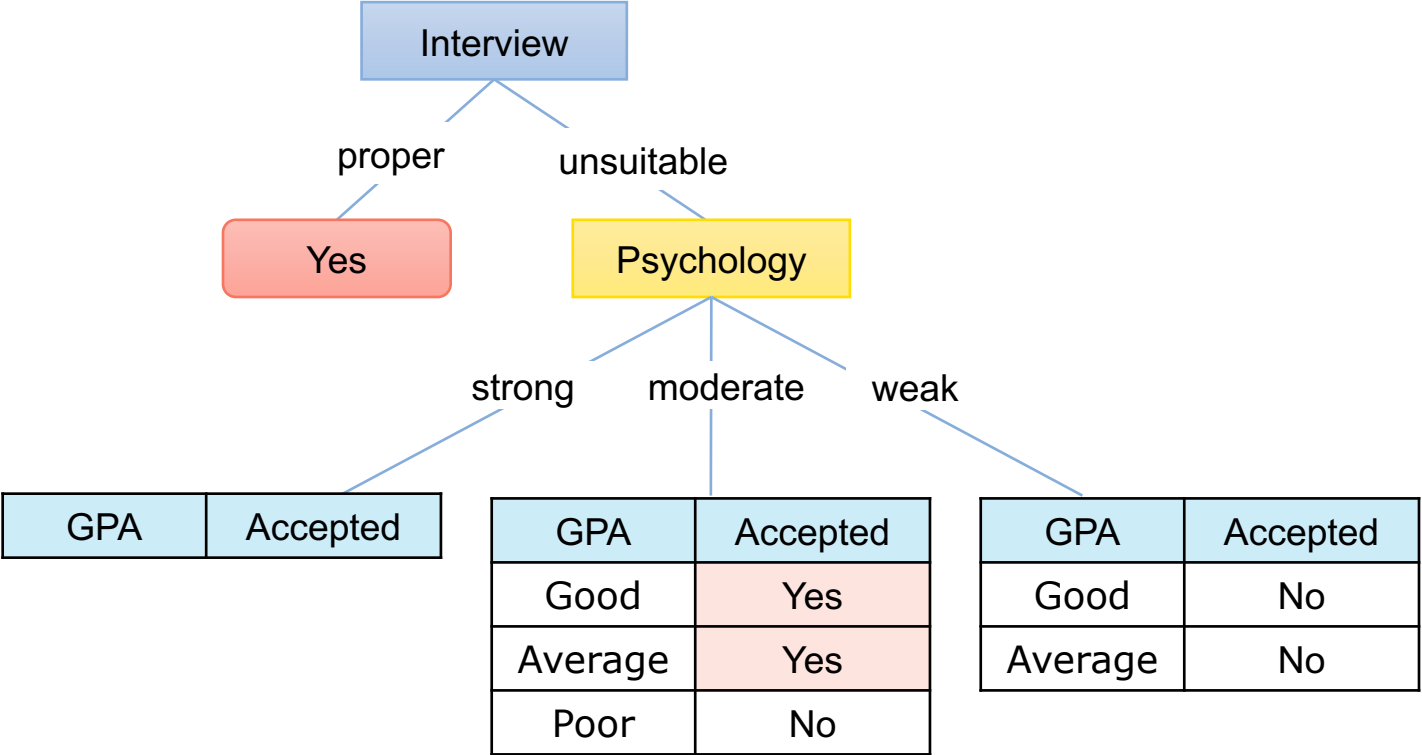
$$Ent(A_m) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9182$$

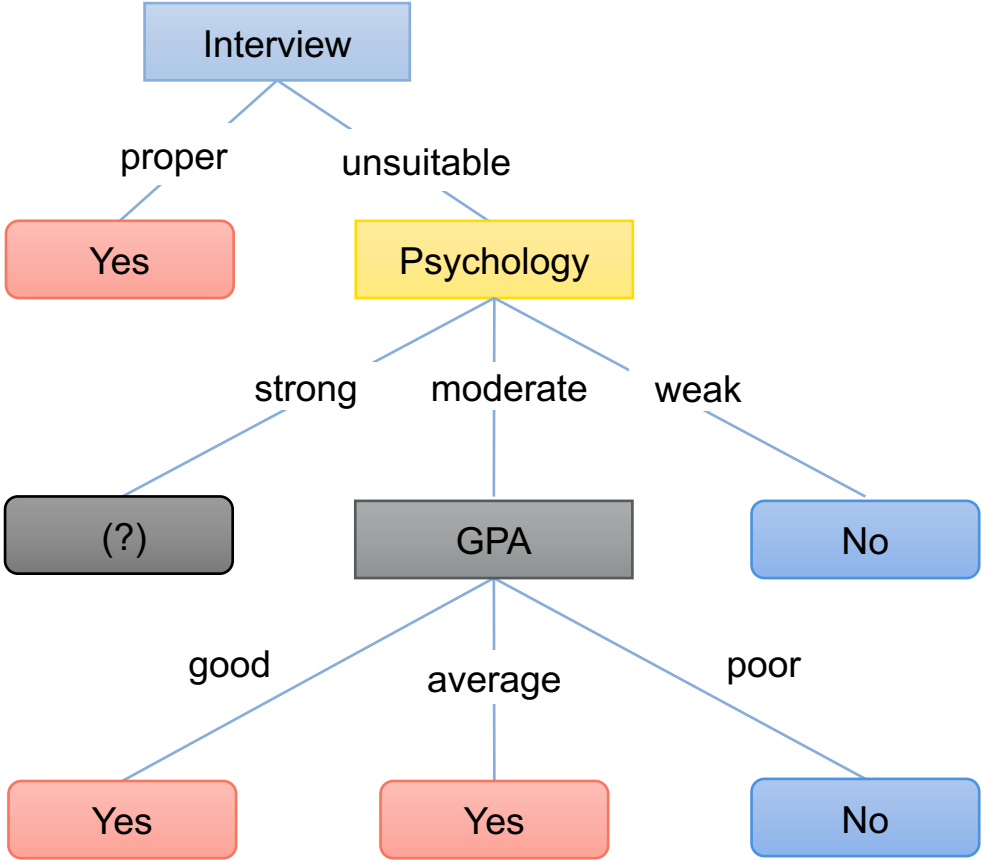
$$Ent(A_w) = 0$$

$$Ent(A_g) = 1 \qquad Ent(A_a) = 1 \qquad Ent(A_p) = 0$$

$$\begin{aligned} Gain(Accepted, GPA) &= Ent(A) - \frac{|A_g|}{|A|}Ent(A_g) - \frac{|A_a|}{|A|}Ent(A_a) - \frac{|A_p|}{|A|}Ent(A_p) \\ &= 0.9709 - \frac{2}{5}1 - \frac{2}{5}1 - \frac{1}{5}0 = 0.1709 \end{aligned}$$

$$\begin{aligned} Gain(Accepted, Psi) &= Ent(A) - \frac{|A_m|}{|A|}Ent(A_m) - \frac{|A_w|}{|A|}Ent(A_w) \\ &= 0.9709 - \frac{3}{5}0.9182 - \frac{2}{5}0 = \mathbf{0.2822} \end{aligned}$$







Mengapa Decision Tree?

48

- Mudah diimplementasikan
- Hipotesis yang dihasilkan mudah dipahami
- Efisien
- Daerah pengambilan dapat diubah menjadi simple dan spesifik.
- Dapat Eliminasi perhitungan-perhitungan yang tidak diperlukan

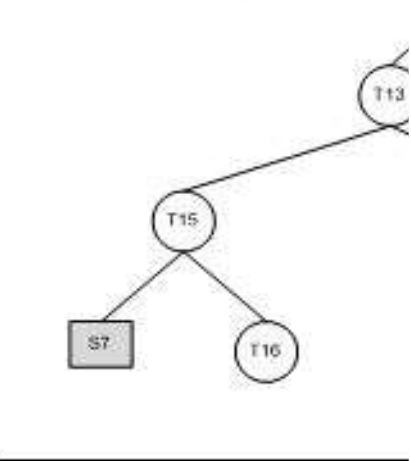
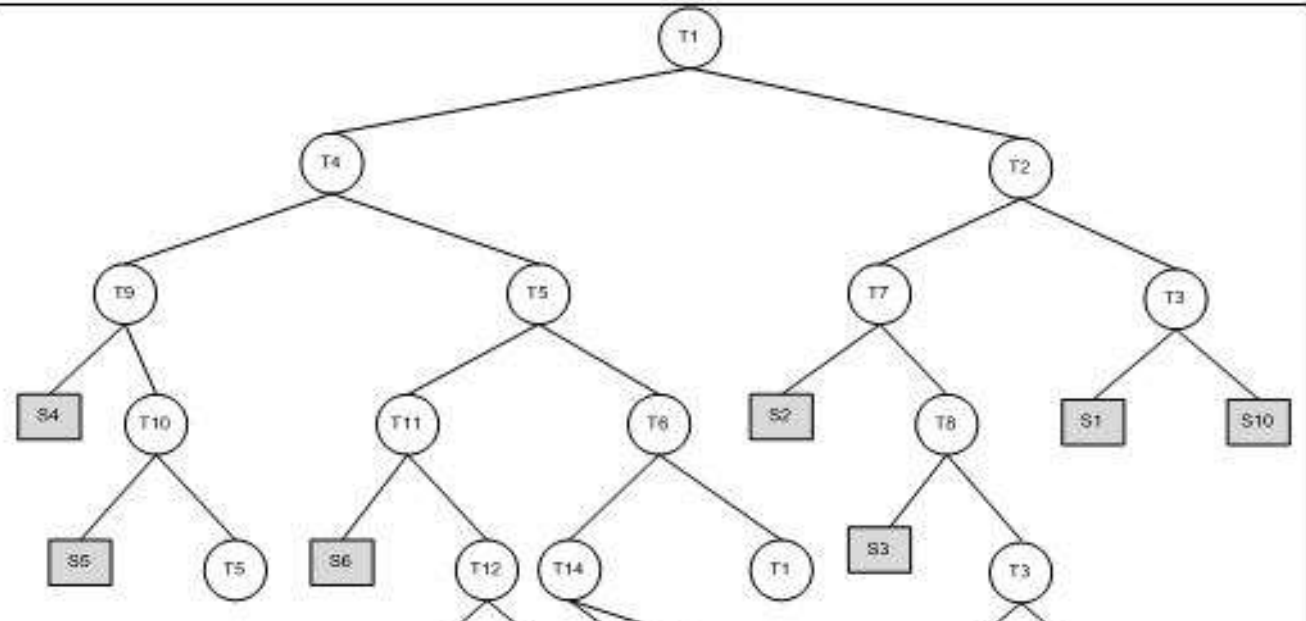
Decision Tree Cocok untuk Masalah:

- Data dalam bentuk atribut-nilai. Kondisi ideal adalah jika isi nilai jumlahnya sedikit. Misalnya: “panas”, “sedang”, “dingin”.
- Output diskrit.

Buat Decision Tree
untuk menentukan
kelas dari suatu buah

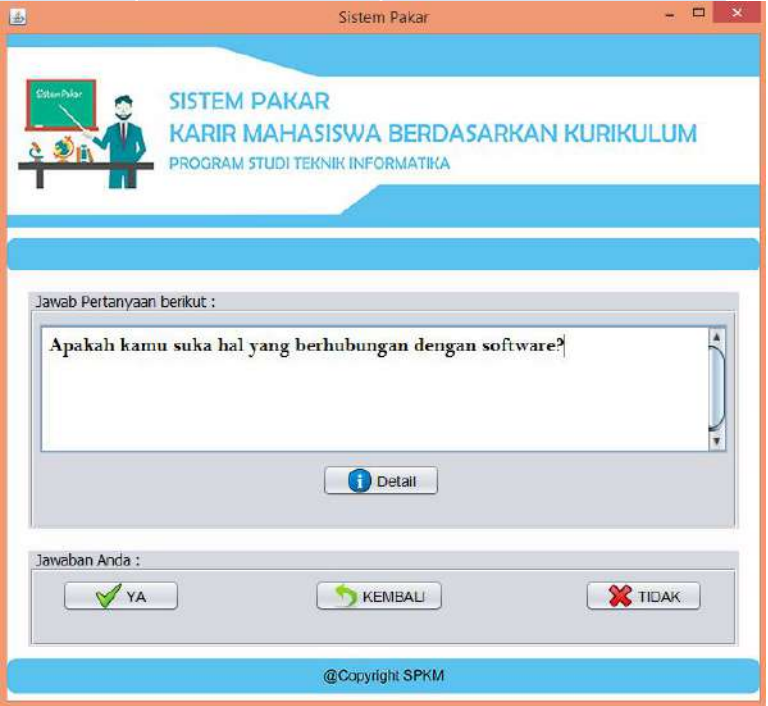
No	Kulit Buah	Warna	Ukuran	Bau	Kelas
1	Kasar	Coklat	Besar	keras	Aman
2	Kasar	Hijau	Besar	keras	Aman
3	Halus	Merah	Besar	Lunak	Berbahaya
4	Kasar	Hijau	Besar	Lunak	Aman
5	Kasar	Merah	Kecil	Keras	Aman
6	Halus	Merah	Kecil	Keras	Aman
7	Halus	Coklat	Kecil	Keras	Aman
8	Kasar	Hijau	Kecil	Lunak	Berbahaya
9	Halus	Hijau	Kecil	Keras	Berbahaya
10	Kasar	Merah	Besar	Keras	Aman
11	Halus	Coklat	Besar	Lunak	Aman
12	Halus	Hijau	Kecil	Keras	Berbahaya
13	Kasar	Merah	Kecil	Lunak	Aman
14	Halus	Merah	Besar	Keras	Berbahaya
15	Halus	Merah	Kecil	Keras	Aman
16	Kasar	Hijau	Kecil	Keras	Berbahaya





ID	Pertanyaan
T1	Apakah kamu suka hal yang berhubungan dengan software?
T2	Apakah kamu suka hal yang berhubungan dengan hardware?
T3	Apakah kamu suka hal yang berhubungan dengan komputer?
T4	Apakah kamu suka hal yang berhubungan dengan coding?
T5	Apakah kamu suka hal yang berhubungan dengan perancangan?
...	

Pertanyaan	Ya	Tidak
T1	T4	T2
T2	T7	T3
T3	S1	S10
T4	T9	T5
...		





ID	Solusi	Detail
S5	Programmer	<p>Programmer adalah orang yang membuat suatu aplikasi untuk kclient atau server baik untuk perusahaan, instansi atupun perorangan.</p> <p>Tugas : Membuat perogram baik aplikasi maupun sistem oprasi dengan menggunakan bahasa pemrograman yang ada.</p> <p>Kualifikasi : Menguasai logika dan algoritma pemrograman seperti HTML, Ajax, CSS, JavaScript, C++, VB, PHP, Java, Ruby dll. Memahami SQL dan menguasai Bahasa Inggris IT.</p>

