Bài 1:

$$Entropy(S) = -\frac{9}{14} \cdot \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \cdot \log_2\left(\frac{5}{14}\right) = 0.94$$
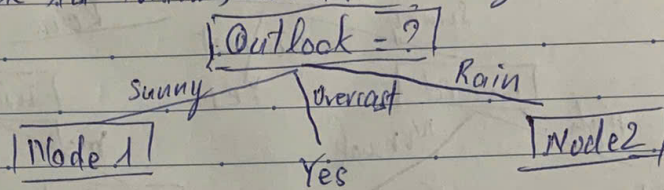
Ⓧ Chọn thuộc tính kttra 1 (gốc)

$$Gain(S, Outlook) = 0.94 - \frac{5}{14} \cdot 0.97 - \frac{4}{14} \cdot 0 - \frac{5}{14} \cdot 0.97 = 0.247$$

$$Gain(S, Temp) = 0.94 - \frac{4}{14} \cdot 1 - \frac{6}{14} \cdot 0.92 - \frac{4}{14} \cdot 0.81 = 0.028$$

$$Gain(S, Humid) = 0.94 - \frac{7}{14} \cdot 0.985 - \frac{7}{14} \cdot 0.592 = 0.1521$$

$$Gain(S, Wind) = 0.94 - \frac{8}{14} \cdot 0.811 - \frac{6}{14} \cdot 0.918 = 0.083$$

$\Rightarrow$ Chọn Outlook là thttính kttra 1 (nút gốc)

[Decision tree diagram:]
Outlook = ?
- Sunny → Node 1
- Overcast → Yes
- Rain → Node 2

Ⓧ Xét Node 1: Entropy $(S_{Sunny}) = 0.97$
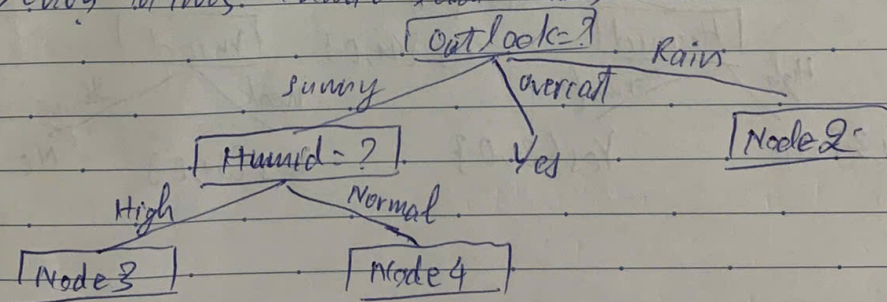
+) Chọn th.tính:

Gain $(S_{Sunny}, Outlook)$

$$Gain(S_{Sunny}, Temp) = 0.97 - \frac{2}{5} \cdot 0 - \frac{2}{5} \cdot 1 - \frac{1}{5} \cdot 0 = 0.57$$

$$Gain(S_{Sunny}, Humid) = 0.97 - \frac{3}{5} \cdot 0 - \frac{2}{5} \cdot 0 = 0.97$$

$$Gain(S_{Sunny}, Wind) = 0.97 - \frac{3}{5} \cdot 0.918 - \frac{2}{5} \cdot 1 = 0.0192$$

$\rightarrow$ Chọn th.tính Humid làm th.tính khảo Node 1

[Decision tree diagram:]
Outlook = ?
- Sunny → Humid = ?
  - High → Node 3
  - Normal → Node 4
- Overcast → Yes
- Rain → Node 2

Ⓧ Xét Node 3: Entropy $(S_{Sunny, High}) = 0$
$\Rightarrow$ dữ liệu thỏa mẫu đều thuộc 1 lớp
$\Rightarrow$ Node 3: nút lá & lá lớp No

Ⓧ Xét Node 4: Entropy $(S_{Sunny, Normal}) = 0$
$\Rightarrow$ thỏa mẫu đều $\in$ 1 lớp
$\Rightarrow$ Node 4: nút lá lớp Yes

[Final decision tree diagram:]
Outlook = ?
- Sunny → Humid
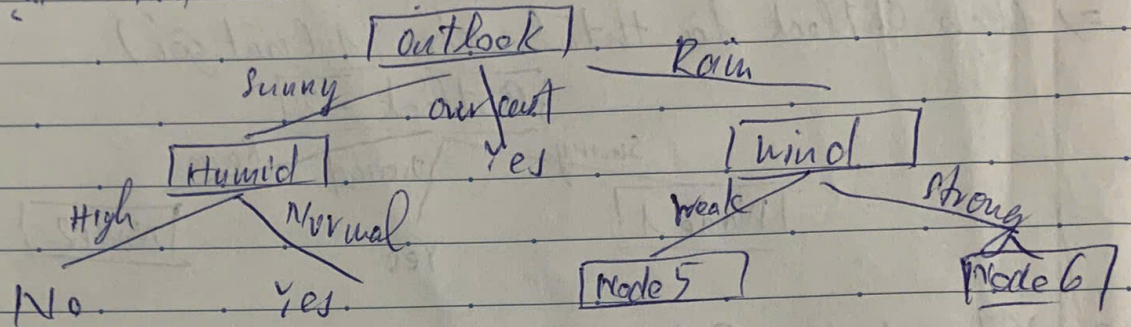 - High → No
  - Normal → Yes
- Overcast → Yes
- Rain → Node 2
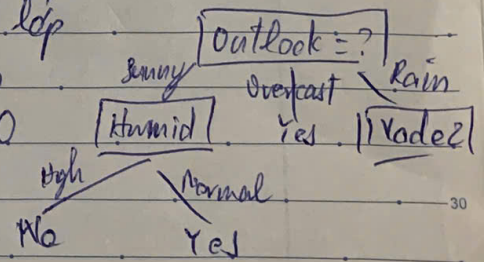
⊛ Xét Node 2: Entropy $(S_{rain}) = 0,97$

Gain $(S_{rain}, Temp) = 0,97 - \frac{0}{5} \cdot - \frac{3}{5} \cdot 0,918 - \frac{2}{5} \cdot 1 = 0,019$

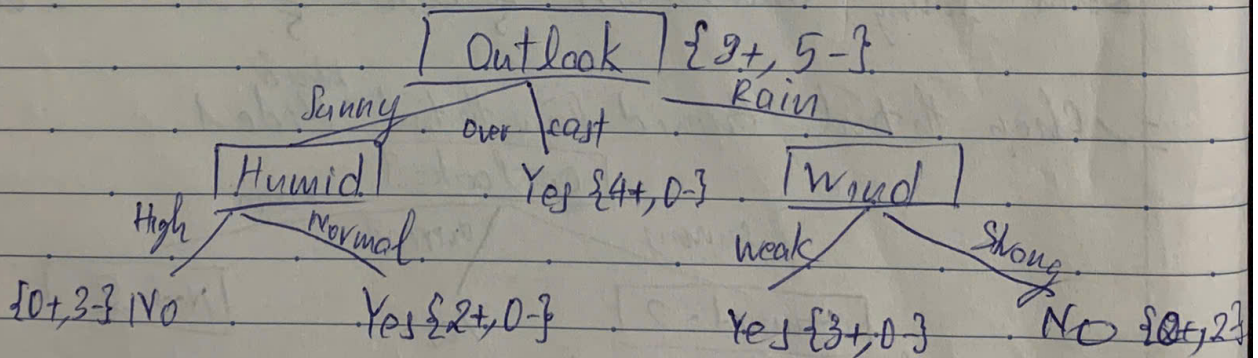Gain $(S_{rain}, Humid) = 0,97 - \frac{2}{5} \cdot 1 - \frac{3}{5} \cdot 0,918 = 0,19$

Gain $(S_{rain}, wind) = 0,97 - \frac{3}{5} \cdot 0 - \frac{2}{5} \cdot 0 = 0,97$

→ chọn wind làm thuộc tính khóa.

[Outlook]

Sunny   over\cast   Rain

[Humid]   Yes   [wind]

High   Normal   weak   strong

No   Yes   [Node 5]   [Node 6]

⊛ Xét Node 5: Entropy $(S_{rain, weak}) = 0$

=> nút là Yes

⊛ Xét Node 6: Entropy $(S_{rain, strong}) = 0$

=> nút là No.

[Outlook] $\{9+, 5-\}$

Sunny   over\cast   Rain

[Humid]   Yes $\{4+, 0-\}$ [wind]

High   Normal   weak   strong

$\{0+, 3-\}$ No   Yes $\{2+, 0-\}$ Yes $\{3+, 0-\}$ No $\{0+, 2-\}$

(*) Bài làm

+, z = {Age = Young, Income = Medium, Student = No, credit_rating = Fair}

+, 2 phân lớp : c_1 : Mua máy tính

c_2 : ko mua máy tính

+, XS tiên nghiệm mỗi phân lớp : $P(c_1) = 9/14$

$P(c_2) = 5/14$

+, XS mỗi thuộc tính xảy ra đối với mỗi phân lớp :

$P(Age = Young | c_1) =$; $P(Age = Medium | c_2) =$

$\dfrac{}{P(Age = Old | c}$

$P(Age = Young | c_1) = 2/9 = 0,222$; $P(Age = Young | c_2) = 3/5 = 0,6$

$P(Age = Medium | c_1) = 4/9 = 0,444$; $P(Age = Medium | c_2) = 0$

$P(Age = Old | c_1) = 3/9 = 0,333$; $P(Age = Old | c_2) = 2/5 = 0,4$

$P(Income = High | c_1) = 2/9 = 0,222$; $P(Income = High | c_2) = 2/5 = 0,4$

$P(Income = Medium | c_1) = 4/9 = 0,444$; $P(Income = Medium | c_2) = 2/5 = 0,4$

$P(Income = Low | c_1) = 3/9 = 0,333$; $P(Income = Low | c_2) = 1/5 = 0,2$

$P(Student = No | c_1) = 3/9 = 0,333$; $P(Student = No | c_2) = 4/5 = 0,8$

$P(Student = Yes | c_1) = 6/9 = 0,667$; $P(Student = Yes | c_2) = 1/5 = 0,2$

$P(Credit\_Rating = Fair | c_1) = 6/9 = 0,667$; $P(Credit\_rating = Fair | c_2) = 2/5 = 0,4$

$P(Credit\_Rating = Excellent | c_1) = 3/9 = 0,333$; $P(Credit\_rating = Excellent | c_2) = 3/5 = 0,6$

+, Tính XS xảy ra giả thuyết (thuộc tính) trong z đối với các phân lớp.

$P(z | c_1)$ = P(Age = Young, Income = Medium, Student = No, Credit_rating = Fair | c_1)

= P(Age = Young | c_1) . P(Income = Medium | c_1) . P(Student = No | c_1) . P(Credit_rating = Fair | c_1)

= $\dfrac{2}{9} \cdot \dfrac{4}{9} \cdot \dfrac{3}{9} \cdot \dfrac{6}{9} = 0,022$

$P(z | c_2)$ = P(Age = Young, Income = Medium, Student = No, Credit_Rating = Fair | c_2)

= P(Age = Young | c_2) . P(Income = Medium | c_2) . P(Student = No | c_2) . P(Credit_Rating = Fair | c_2)

= $\dfrac{3}{5} \cdot \dfrac{2}{5} \cdot \dfrac{4}{5} \cdot \dfrac{2}{10} = 0,077$

+) Xác định lớp z có thể rơi vào nhất

xv phân lớp c_1 = $P(c_1) \cdot P(z|c_1)$ = $9/14 \cdot 0,022 = 0,014$

xv phân lớp c_2 = $P(c_2) \cdot P(z|c_2)$ = $5/14 \cdot 0,077 = 0,028$

$\Rightarrow$ Âm Người có thuộc tính đặc như z khả năng sử
tập của máy tính