นายธีรนันท์ ทิพย์พิมานพงษ์
6430205165-5 SC-SF

" HW 2 "

Attribute Selection with Information Gain.

(yes, no)

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
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
>40	low	yes	excellent	no

|—— Feature ——| Class

age: <=30 (2,3) 5; 31...40 (4,0) 4; >40 (3,2) 5

income: high (2,2) 4; med (4,2) 6; low (3,1) 4

student: yes (6,1) 7; No (3,5) 8

credit: excellent (3,3) 6; fair (6,2) 8

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

$$= I(9,5) \quad (\text{yes}, \text{no})$$

$$= -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$$

$$= 0.940 \#$$

– Expected info (entropy) ของทั้งหมด

- คำนวณ Feature จาก $Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$ → คำนวณเหมือน Class

Expected info แต่ละตัวเพื่อหา Root node

$$Info_{age}(D) = \frac{5}{14}I(2,3) + \frac{4}{14}I(4,0) + \frac{5}{14}I(3,2) = 0.6935 \#$$

$$Info_{income}(D) = \frac{4}{14}I(2,2) + \frac{6}{14}I(4,2) + \frac{4}{14}I(3,1) = 0.9109 \#$$

$$Info_{student}(D) = \frac{7}{14}I(6,1) + \frac{7}{14}I(3,4) = 0.7884 \#$$

$$Info_{credit}(D) = \frac{8}{14}I(6,2) + \frac{6}{14}I(3,3) = 0.8921 \#$$

Gain เท่ากับ Root node

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

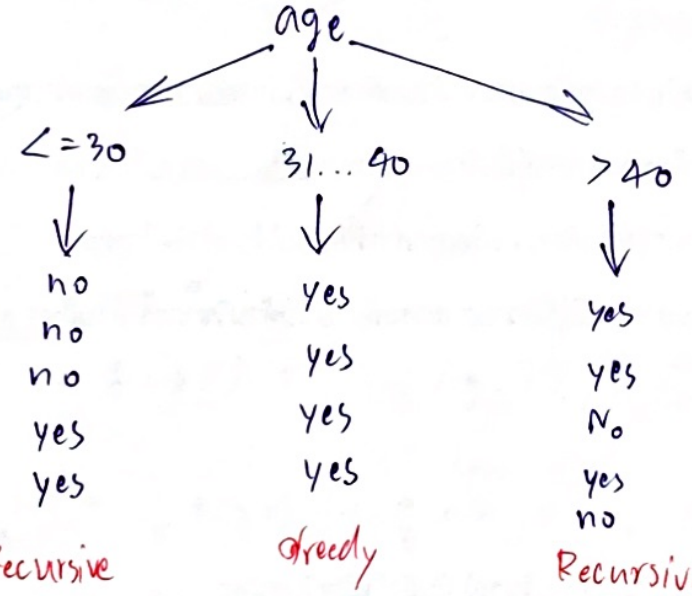
$$= 0.940 - \text{Info}_A(D)$$

Gain (age) = 0.940 − 0.6935 = **0.2465** → เลือก age เป็น root node

Gain (income) = 0.940 − 0.9109 = 0.0291

Gain (student) = 0.940 − 0.7884 = 0.1516

Gain (credit_rating) = 0.940 − 0.8921 = 0.0479



Recursive age <=30

$$\text{Info}(D) = I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.9710$$

$$\text{Info}_{\text{income}}(D) = \frac{2}{5}I(0,2)_{\text{high}} + \frac{2}{5}I(1,1)_{\text{med}} + \frac{1}{5}I(1,0)_{\text{low}} = 0.4$$

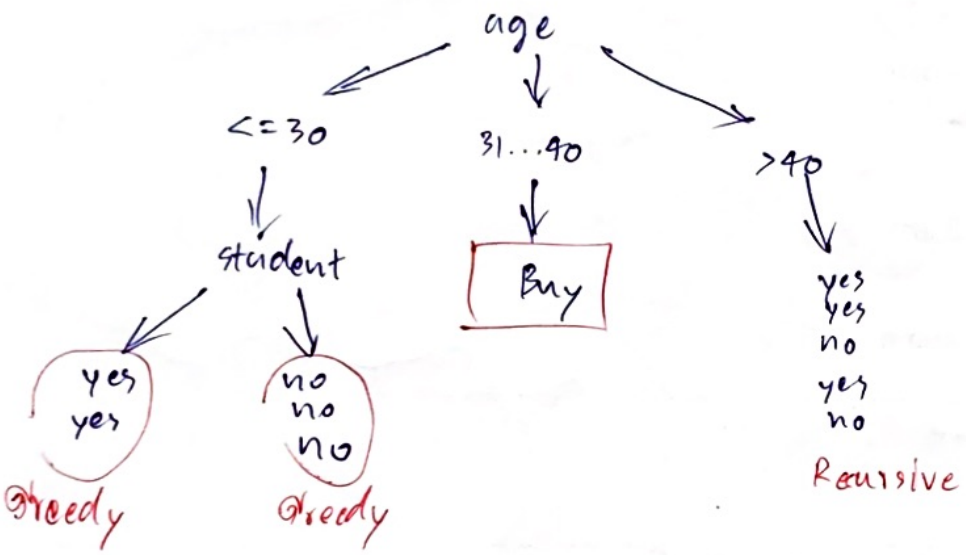
$$\text{Info}_{\text{student}}(D) = \frac{2}{5}I(1,0)_{\text{yes}} + \frac{3}{5}I(0,3)_{\text{no}} = 0$$

$$\text{Info}_{\text{Credit\_rating}}(D) = \frac{2}{5}I(1,1)_{\text{excellent}} + \frac{3}{5}I(1,2)_{\text{fair}} = 0.9509$$

Gain (income) = 0.9710 − 0.4 = 0.5710

Gain (student) = 0.9710 − 0 = 0.9710 → เลือก student เป็น root node

Gain (credit_rating) = 0.9710 − 0.9509 = 0.0201



Recursive age >40

$$Info(D) = I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.9710$$

$$Info_{income}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1) = 0.9509$$ (med, low)

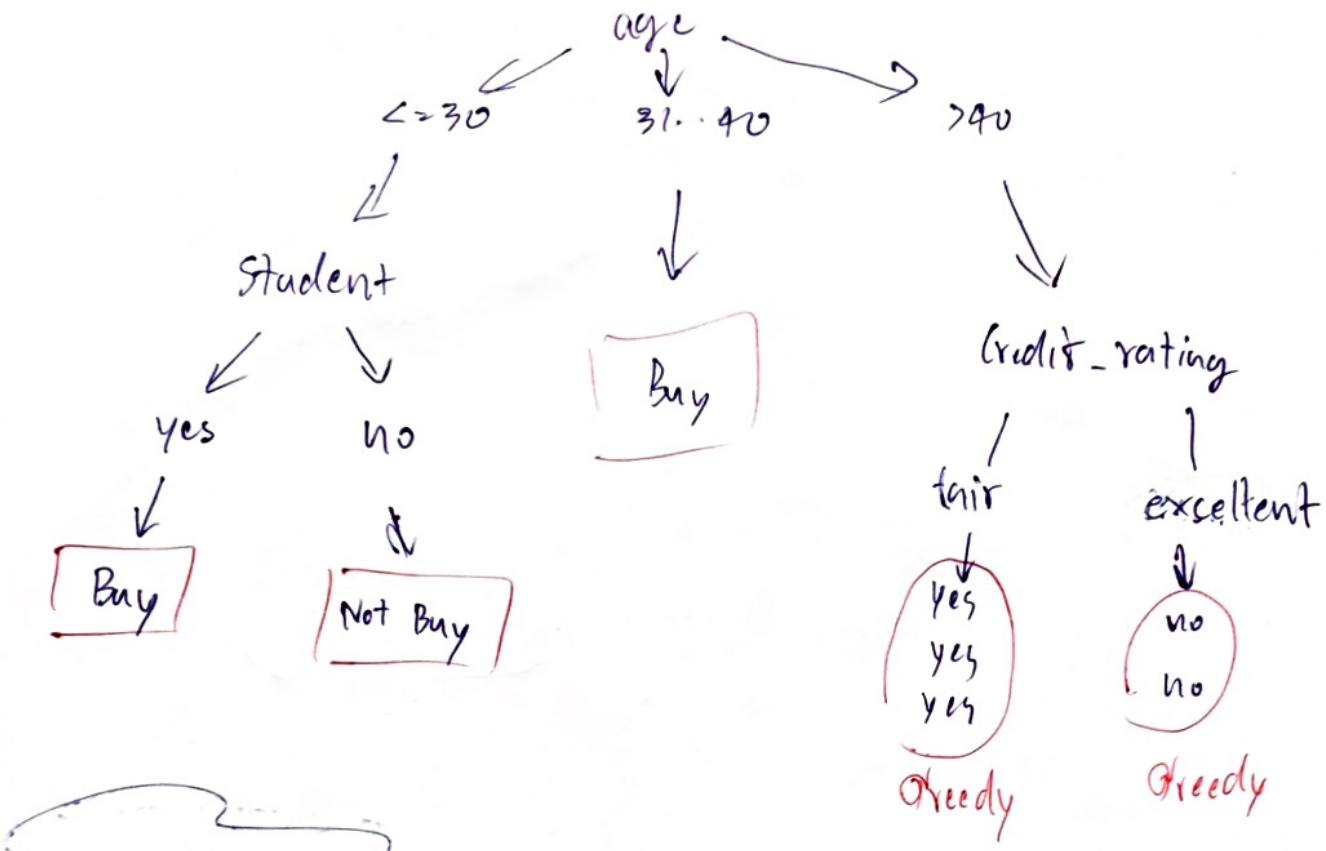
$$Info_{student}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1) = 0.9509$$

$$Info_{credit\_rating}(D) = \frac{2}{5}I(0,2) + \frac{3}{5}I(3,0) = 0$$

Gain (income) = 0.9710 − 0.9509 = 0.0201

Gain (student) = 0.9710 − 0.9509 = 0.0201

Gain (credit_rating) = 0.9710 − 0 = 0.9710 → เลือก credit_rating เป็น root node



Decision Tree