

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

features = X

class = Y

คำนวณ class

$$\begin{aligned}
 \text{Info}(D) &= \sum_{i=1}^m p_i \log_2(p_i) \\
 &= -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) \\
 &= 0.41 + 0.53 \\
 \text{Info}(D) &= 0.940
 \end{aligned}$$

คำนวณ Feature Info<sub>A</sub>(D)

$$\text{Info}_A(D) = \sum_{i=1}^n \frac{|D_i|}{|D|} \times \text{Info}(D_i)$$

$$\begin{aligned}
 \text{Info}_{\text{age}}(D) &= \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) \\
 &= \frac{5}{14} \left[ -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right] + \frac{4}{14} \left[ -\frac{4}{4} \log_2\left(\frac{4}{4}\right) - \frac{0}{4} \log_2\left(\frac{0}{4}\right) \right] + \\
 &\quad \frac{5}{14} \left[ -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \right] \\
 &= 0.34676 + 0.34676 = 0.69352
 \end{aligned}$$

$$\therefore \text{Info}_{\text{age}}(D) = 0.694$$

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) \\
 &= \frac{4}{14} \left[ -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) \right] + \frac{6}{14} \left[ -\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) \right] + \frac{4}{14} \left[ -\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) \right] \\
 &= 0.2857 + 0.3935 + 0.2317 = 0.9109
 \end{aligned}$$

$$\therefore \text{Info}_{\text{income}}(D) = 0.911 \neq$$

$$\begin{aligned}
 \text{Info}_{\text{Credit}}(D) &= \frac{6}{14} I(3,3) + \frac{8}{14} I(6,2) \\
 &= \frac{6}{14} \left[ -\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right] + \frac{8}{14} \left[ -\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) \right] \\
 &= \frac{6}{14} [0.5 + 0.5] + \frac{8}{14} [0.3113 + 0.5] \\
 &= \left( \frac{6}{14} \times 1 \right) + \left( \frac{8}{14} \times 0.8113 \right) \\
 &= 0.4285 + 0.4636 = 0.8921
 \end{aligned}$$

$$\therefore \text{Info}_{\text{Credit}}(D) = 0.892 \neq$$

$$\begin{aligned}
 \text{Info}_{\text{student}}(D) &= \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) \\
 &= \frac{7}{14} \left[ -\frac{1}{7} \log_2\left(\frac{1}{7}\right) - \frac{6}{7} \log_2\left(\frac{6}{7}\right) \right] + \frac{7}{14} \left[ -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) \right] \\
 &= \frac{7}{14} [0.1906 + 0.4010] + \frac{7}{14} [0.5238 + 0.4615] \\
 &= 0.2958 + 0.4925 = 0.7883
 \end{aligned}$$

$$\therefore \text{Info}_{\text{student}}(D) = 0.7883 \neq$$

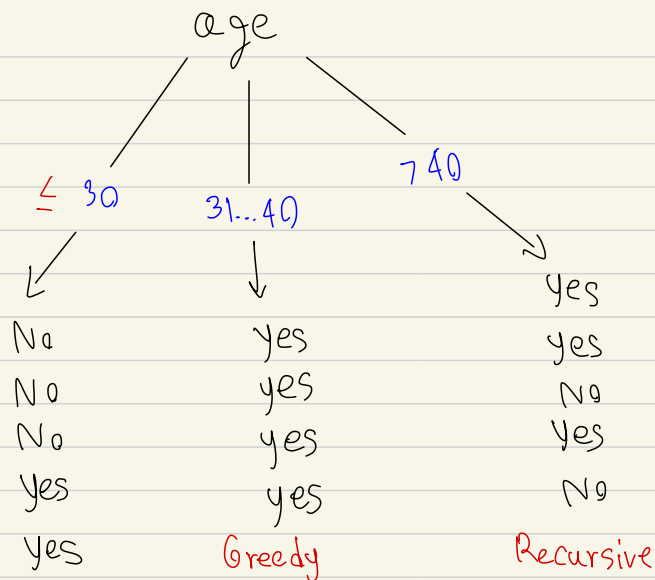
$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

$$\text{Gain}(\text{age}) = 0.940 - 0.694 = 0.246 \rightarrow \text{Gain สูงที่สุด} \text{ เลือกเป็น root node}$$

$$\text{Gain}(\text{income}) = 0.940 - 0.911 = 0.029$$

$$\text{Gain}(\text{student}) = 0.940 - 0.788 = 0.152$$

$$\text{Gain}(\text{Credit-rating}) = 0.940 - 0.892 = 0.048$$



Recursive age ≤ 30

Recursive

X

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

Entropy Class

$$\text{Info}(D) = I(2, 3)$$

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

$$= 0.5288 + 0.4422$$

$$\therefore \text{Info}(D) = 0.9710 \neq$$

Entropy Feature

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{2}{5} I(0, 2) + \frac{2}{5} I(1, 1) + \frac{1}{5} I(1, 0) \\
 &= \frac{2}{5} \left[ -\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) \right] + \frac{2}{5} \left[ -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \\
 &\quad + \frac{1}{5} \left[ -\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right]
 \end{aligned}$$

$$= \frac{2}{5} \left( \cancel{\ln \frac{1}{2}} \right) + \frac{2}{5} (0.5 + 0.5) + \frac{1}{5} \left( \cancel{\ln \frac{1}{5}} \right)$$

$$= \frac{2}{5} \times 1$$

$$\therefore \text{Info}_{\text{income}}(D) = 0.4 \quad \times$$

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

$$= \frac{2}{5} \left[ -\frac{2}{2} \log_2 \left( \frac{2}{2} \right) - \frac{0}{2} \log_2 \left( \frac{0}{2} \right) \right] + \frac{3}{5} \left[ -\frac{0}{3} \log_2 \left( \frac{0}{3} \right) - \frac{3}{3} \log_2 \left( \frac{3}{3} \right) \right]$$

$$= \frac{2}{5} \left[ \cancel{\ln \frac{1}{2}} \right] + \frac{3}{5} \left[ \cancel{\ln \frac{1}{3}} \right]$$

$$= 0 + 0$$

$$\therefore \text{Info}_{\text{student}}(D) = 0 \quad \times$$

$$\text{Info}_{\text{credit}}(D) = \frac{2}{5} I(1,1) + \frac{3}{5} I(1,2)$$

$$= \frac{2}{5} \left[ -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right] + \frac{3}{5} \left[ -\frac{1}{3} \log_2 \left( \frac{1}{3} \right) - \frac{2}{3} \log_2 \left( \frac{2}{3} \right) \right]$$

$$= \frac{2}{5} [0.5 + 0.5] + \frac{3}{5} [0.5283 + 0.3899]$$

$$= \left[ \frac{2}{5} \times 1 \right] + \left[ \frac{3}{5} \times 0.9182 \right]$$

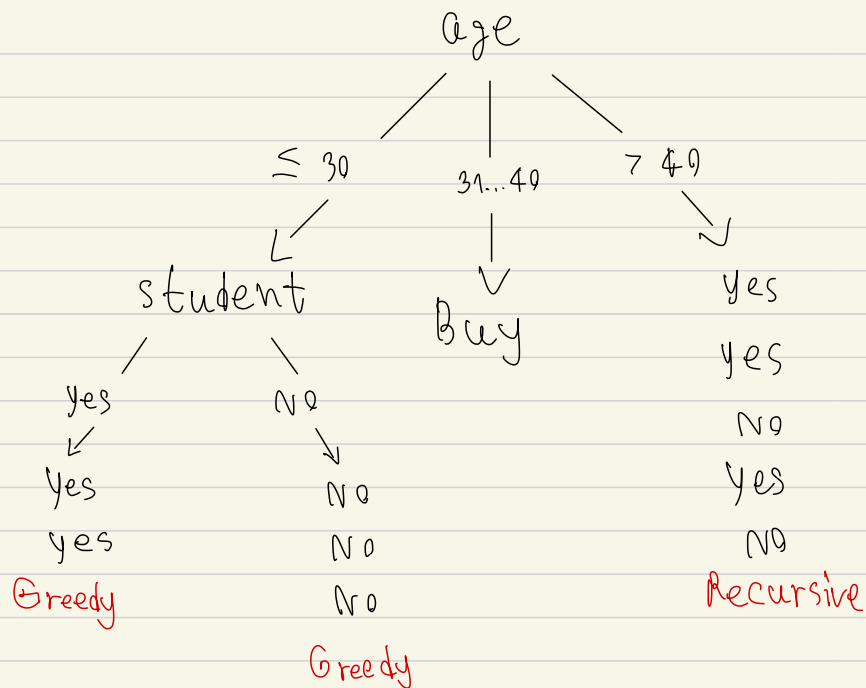
$$= 0.4 + 0.5509$$

$$\therefore \text{Info}_{\text{credit}}(D) = 0.9509 \quad \times$$

$$\text{Gain}(\text{income}) = 0.9710 - 0.4 = 0.5710$$

$$\text{Gain}(\text{student}) = 0.9710 - 0 = 0.9710 \rightarrow \text{Gain มากที่สุด เลือก}$$

$$\text{Gain}(\text{credit-rating}) = 0.9710 - 0.9509 = 0.0201$$



Recursive age > 40

X

X

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

Answer Class

$$Info(D) = I(3,2)$$

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

$$= 0.4422 + 0.5288$$

$$\therefore Info(D) = 0.971 \neq$$

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1) \\
 &= \frac{3}{5} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] + \frac{2}{5} \left[ -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right] \\
 &= \frac{3}{5} [0.3899 + 0.5283] + \frac{2}{5} [0.5 + 0.5] \\
 &= 0.5509 + 0.4 \\
 \text{Info}_{\text{income}}(D) &= 0.9509
 \end{aligned}$$

$$\begin{aligned}
 \text{Info}_{\text{credit}}(D) &= \frac{2}{5} I(0,2) + \frac{3}{5} I(3,0) \\
 &= \frac{2}{5} \left[ -\frac{0}{2} \log_2 \left( \frac{0}{2} \right) - \frac{2}{2} \log_2 \left( \frac{2}{2} \right) \right] + \frac{3}{5} \left[ -\frac{3}{3} \log_2 \left( \frac{3}{3} \right) - \frac{0}{3} \log_2 \left( \frac{0}{3} \right) \right] \\
 &= \frac{2}{5} [0] + \frac{3}{5} [0] \\
 \therefore \text{Info}_{\text{credit}}(D) &= 0
 \end{aligned}$$

$$\text{Gain}(\text{income}) = 0.9710 - 0.9509 = 0.0201$$

$$\text{Gain}(\text{credit\_rating}) = 0.9710 - 0 = 0.9710 \rightarrow \text{Gain max}$$

