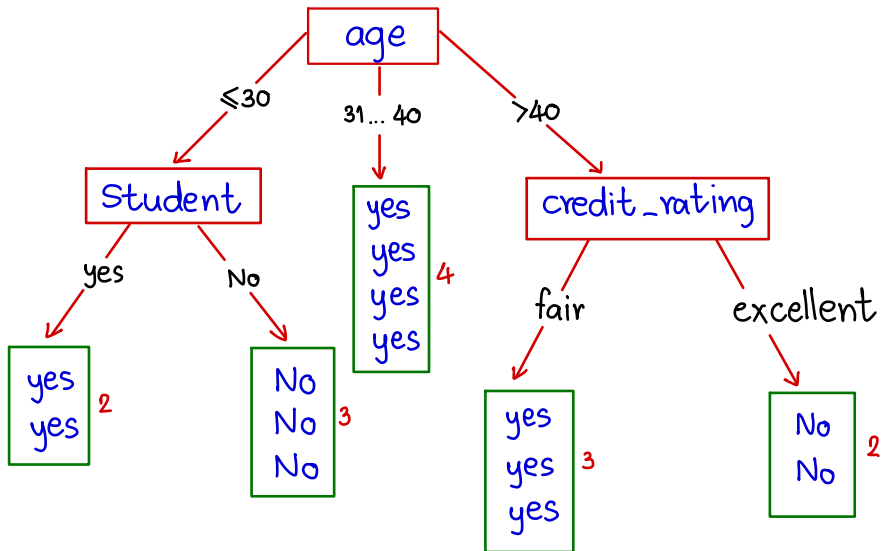


age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes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



min simple = 4

$$\text{Info}(D) = I(\overset{\text{yes}}{9}, \overset{\text{no}}{5}) = -\frac{9}{14} \log_2 \left(\frac{9}{14} \right) - \frac{5}{14} \log_2 \left(\frac{5}{14} \right) = 0.940$$

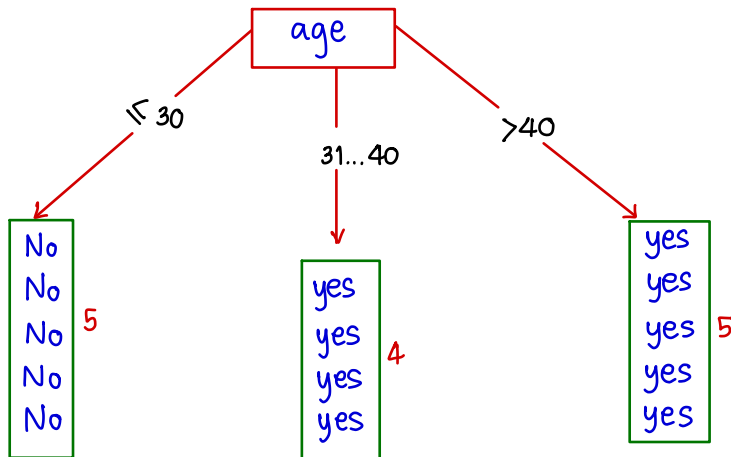
$$\begin{aligned} \text{Info}_{\text{age}}(D) &= \overset{\leq 30}{\frac{5}{14}} I(2, 3) + \overset{31-40}{\frac{4}{14}} I(4, 0) + \overset{>40}{\frac{5}{14}} I(3, 2) \\ &= \frac{5}{14} \left[-\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{5} \right) \right] + \\ &\quad \frac{4}{14} \left[-\frac{4}{4} \log_2 \left(\frac{4}{4} \right) - \frac{0}{4} \log_2 \left(\frac{0}{4} \right) \right] + \\ &\quad \frac{5}{14} \left[-\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right] \\ &= 0.694 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \overset{\text{high}}{\frac{4}{14}} I(2, 2) + \overset{\text{medium}}{\frac{6}{14}} I(4, 2) + \overset{\text{low}}{\frac{4}{14}} I(3, 1) \\ &= \frac{4}{14} \left[-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right] + \\ &\quad \frac{6}{14} \left[-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right] + \\ &\quad \frac{4}{14} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] \\ &= 0.911 \end{aligned}$$

$$\begin{aligned}
 \text{Info}_{\text{student}}(D) &= \frac{4}{14} \overset{\text{yes}}{I}(6,1) + \frac{4}{14} \overset{\text{no}}{I}(3,4) \\
 &= \frac{4}{14} \left[-\frac{6}{4} \log_2 \left(\frac{6}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] + \\
 &\quad \frac{4}{14} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{4}{4} \log_2 \left(\frac{4}{4} \right) \right] \\
 &= 0.788
 \end{aligned}$$

$$\begin{aligned}
 \text{Info}_{\text{credit-rating}}(D) &= \frac{6}{14} \overset{\text{excellent}}{I}(3,3) + \frac{8}{14} \overset{\text{fair}}{I}(6,2) \\
 &= \frac{6}{14} \left[-\frac{3}{6} \log_2 \left(\frac{3}{6} \right) - \frac{3}{6} \log_2 \left(\frac{3}{6} \right) \right] + \\
 &\quad \frac{8}{14} \left[-\frac{6}{8} \log_2 \left(\frac{6}{8} \right) - \frac{2}{8} \log_2 \left(\frac{2}{8} \right) \right] \\
 &= 0.892
 \end{aligned}$$

Gain age = 0.940 - 0.694 = 0.246
 Gain income = 0.940 - 0.911 = 0.029
 Gain student = 0.940 - 0.788 = 0.152
 Gain credit-rating = 0.940 - 0.892 = 0.048



Max_feature = 2

เลือก 2 ฟีเชอร์ จาก age, income, student, credit_rating
คือ income และ credit_rating

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

income

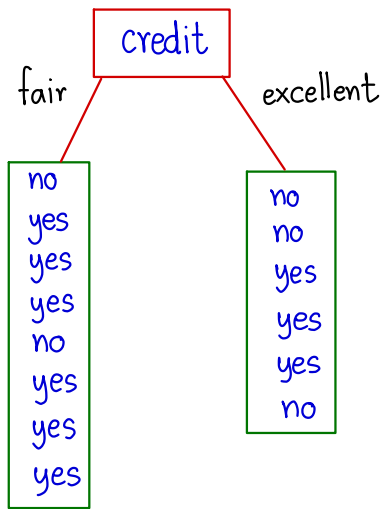
$$\begin{aligned}\text{Info}_{\text{income}}(D) &= \frac{4}{14} I^{\text{high}}(2,2) + \frac{6}{14} I^{\text{mediem}}(4,2) + \frac{4}{14} I^{\text{low}}(3,1) \\&= \frac{4}{14} \left[-\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) \right] + \frac{6}{14} \left[-\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) \right] \\&\quad + \frac{4}{14} \left[-\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) \right] \\&= 0.911\end{aligned}$$

credit_rating

$$\begin{aligned}\text{Info}_{\text{credit_rating}}(D) &= \frac{6}{14} I^{\text{excellent}}(3,3) + \frac{8}{14} I^{\text{fair}}(6,2) \\&= \frac{6}{14} \left[-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right] + \frac{8}{14} \left[-\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) \right] \\&= 0.892\end{aligned}$$

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

$$\text{Gain}(\text{credit_rating}) = \text{Info}(D) - \text{Info}_{\text{credit_rating}}(D) = 0.940 - 0.892 = 0.048$$



ให้ credit-rating ใหม่

credit_rating = fair

$$\text{Info}(D) = I(6,2) = -\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) = 0.811$$

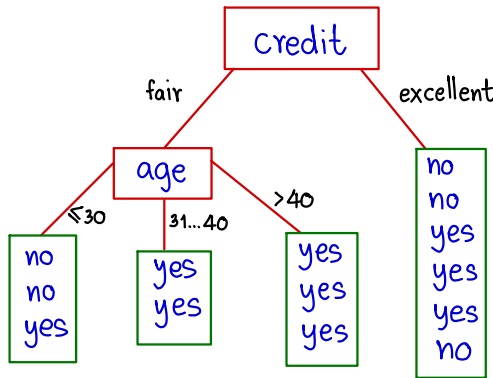
สุ่ม 2 ตัวจาก age, income, student จะได้ income กับ age

$$\begin{aligned}
 \text{info}_{\text{income}}(D) &= \frac{3}{8} \overset{\text{high}}{I(2,1)} + \frac{3}{8} \overset{\text{mediem}}{I(2,1)} + \frac{2}{8} \overset{\text{low}}{I(2,0)} \\
 &= \frac{3}{8} \left[-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \right] + \frac{3}{8} \left[-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \right] \\
 &\quad + \frac{2}{8} \left[-\frac{2}{2} \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \log_2\left(\frac{0}{2}\right) \right] \\
 &= 0.396
 \end{aligned}$$

$$\begin{aligned}
 \text{info}_{\text{age}}(D) &= \frac{3}{8} \overset{\leq 30}{I(1,2)} + \frac{2}{8} \overset{31 \dots 40}{I(2,0)} + \frac{3}{8} \overset{> 40}{I(3,0)} \\
 &= \frac{3}{8} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right] + \frac{2}{8} \left[-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) - \frac{0}{2} \log_2 \left(\frac{0}{2} \right) \right] \\
 &\quad + \frac{3}{8} \left[-\frac{3}{3} \log_2 \left(\frac{3}{3} \right) - \frac{0}{3} \log_2 \left(\frac{0}{3} \right) \right] \\
 &= 0.344
 \end{aligned}$$

$$\text{Gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.811 - 0.396 = 0.415$$

$$\text{Gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.811 - 0.344 = 0.467$$



กรณี age ≤ 30

ใช่ student, income ไม่

credit = fair, age ≤ 30

$$\text{Info}(D) = I(1,2) = -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) = 0.918$$

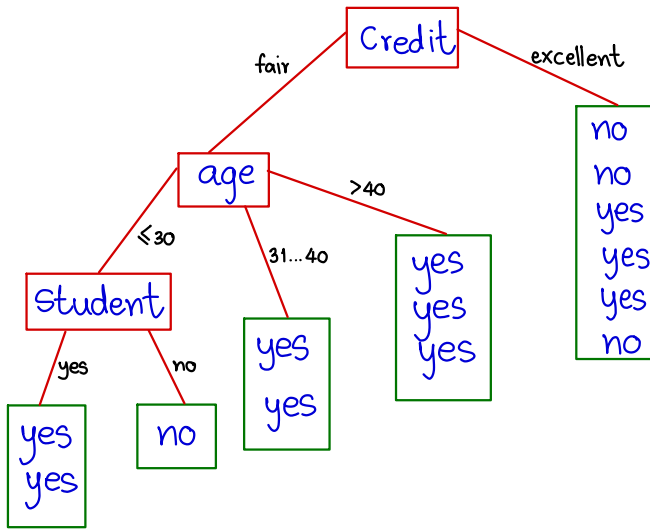
$$\begin{aligned}\text{Info}_{\text{student}}(D) &= \frac{2}{3} I(0,2) + \frac{1}{3} I(1,0) \\ &= \frac{2}{3} \left[-\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) \right] + \frac{1}{3} \left[-\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right] \\ &= 0\end{aligned}$$

$$\begin{aligned}\text{Info}_{\text{income}}(D) &= \frac{1}{3} I(0,1) + \frac{1}{3} I(0,1) + \frac{1}{3} I(1,0) \\ &= \frac{1}{3} \left[-\frac{0}{1} \log_2\left(\frac{0}{1}\right) - \frac{1}{1} \log_2\left(\frac{1}{1}\right) \right] + \frac{1}{3} \left[-\frac{0}{1} \log_2\left(\frac{0}{1}\right) - \frac{1}{1} \log_2\left(\frac{1}{1}\right) \right] \\ &\quad + \frac{1}{3} \left[-\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right] \\ &= 0\end{aligned}$$

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.918 - 0 = 0.918$$

$$\text{Gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.918 - 0 = 0.918$$

สามารถเลือกใช่หรือไม่ก็ได้



credit_rating = excellent

$$\text{Info}(D) = I(3,3) = -\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) = 1$$

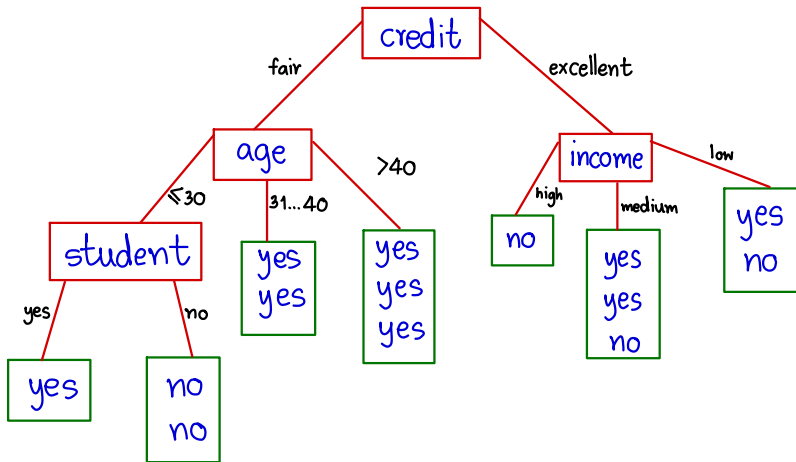
ข้อมูล 2 ตัวคือ age, income, student จะได้ income กับ student

$$\begin{aligned} \text{Info}_{\text{student}}(D) &= \frac{3}{6} I_{\text{yes}}(1,2) + \frac{3}{6} I_{\text{no}}(1,2) \\ &= \frac{3}{6} \left[-\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \right] + \frac{3}{6} \left[-\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \right] \\ &= 0.918 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \frac{1}{6} I_{\text{high}}(0,1) + \frac{3}{6} I_{\text{medium}}(2,1) + \frac{2}{6} I_{\text{low}}(1,1) \\ &= \frac{1}{6} \left[-\frac{0}{1} \log_2\left(\frac{0}{1}\right) - \frac{1}{1} \log_2\left(\frac{1}{1}\right) \right] + \frac{3}{6} \left[-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \right] \\ &\quad + \frac{2}{6} \left[-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \\ &= 0.759 \end{aligned}$$

$$\text{Gain (student)} = \text{Info (D)} - \text{Info}_{\text{student}}(\text{D}) = 1 - 0.918 = 0.082$$

$$\text{Gain (income)} = \text{Info (D)} - \text{Info}_{\text{income}}(\text{D}) = 1 - 0.759 = 0.241$$



အဆိုပါ medium

ဤ student, age များ

credit = excellent , Income = medium

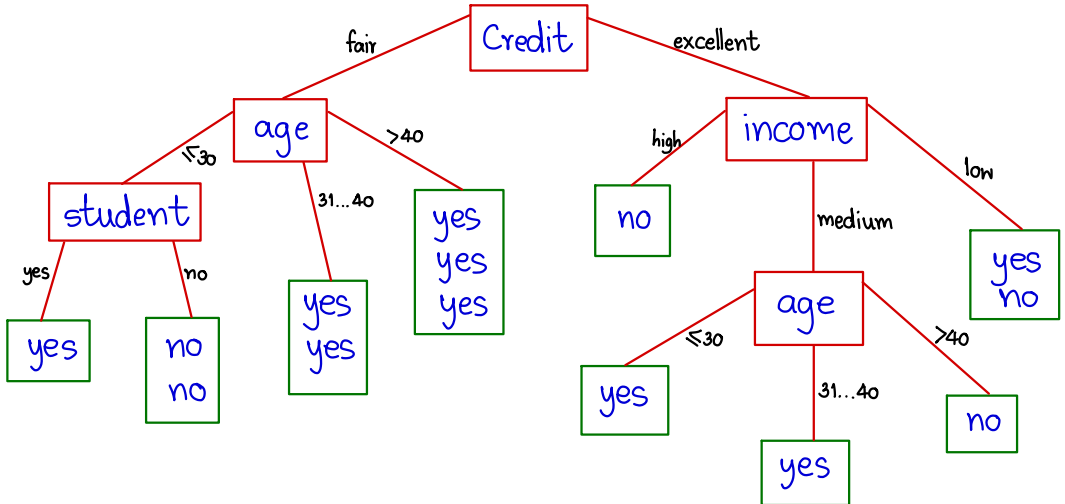
$$\text{Info (D)} = I(2,1) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$$

$$\begin{aligned} \text{Info}_{\text{student}}(\text{D}) &= \frac{1}{3} I(1,0) + \frac{2}{3} I(1,1) \\ &= \frac{1}{3} \left[-\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right] + \frac{2}{3} \left[-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \\ &= 0.6 \end{aligned}$$

$$\begin{aligned}
 \text{Info}_{\text{age}}(D) &= \frac{1}{3} I(1,0) + \frac{1}{3} I(1,0) + \frac{1}{3} I(0,1) \\
 &= \frac{1}{3} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right] + \frac{1}{3} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right] \\
 &\quad + \frac{1}{3} \left[-\frac{0}{1} \log_2 \left(\frac{0}{1} \right) - \frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right] \\
 &= 0
 \end{aligned}$$

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.918 - 0.6 = 0.318$$

$$\text{Gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.918 - 0 = 0.918$$



กรณี Income = low

ถ้า student, age ไม่

Credit = excellent, Income = low

$$\text{Info}(D) = I(1,1) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$\text{Info}_{\text{age}}(D) = \frac{1}{2} I^{31...40}(1,0) + \frac{1}{2} I^{>40}(0,1)$$

$$= \frac{1}{2} \left[-\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right] + \frac{1}{2} \left[-\frac{0}{1} \log_2\left(\frac{0}{1}\right) - \frac{1}{1} \log_2\left(\frac{1}{1}\right) \right]$$
$$= 0$$

$$\text{Info}_{\text{student}}(D) = \frac{2}{2} I^{\text{yes}}(1,1)$$

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

$$\text{Gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 1 - 0 = 1$$

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 1 - 1 = 0$$

final

