Class (buys)

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

$$= I(9,5)$$

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

$$= -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14}$$

$$= -\frac{9}{14}(-0.637) - \frac{5}{14}(-1.485)$$

$$= 0.940 \ \#$$

Feature

$$Info_{age}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

$$= \frac{5}{14}\left[-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] + \frac{4}{14}\left[-\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right)\right] + \frac{5}{14}\left[-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right]$$

(0 → หาค่าไม่ได้)

$$= \frac{5}{14}(0.529 + 0.442) + \frac{4}{14}(0 + \text{หาค่าไม่ได้}) + \frac{5}{14}(0.442 + 0.529)$$

$$= \frac{5}{14}(0.971) + \frac{5}{14}(0.971)$$

$$= 0.347 + 0.347$$

$$= 0.694 \ \#$$

$$Info_{income}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

$$= \frac{4}{14}I(2,2) + \frac{6}{14}I(4,2) + \frac{4}{14}I(3,1)$$

$$= \frac{4}{14}\left[-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right] + \frac{6}{14}\left[-\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right] + \frac{4}{14}\left[-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right]$$

$$= \frac{4}{14}(0.5 + 0.5) + \frac{6}{14}(0.390 + 0.528) + \frac{4}{14}(0.311 + 0.5)$$

$$= \frac{4}{14} + \frac{6}{7}(0.918) + \frac{4}{14}(0.811)$$

$$= 0.286 + 0.394 + 0.232$$

$$= 0.912 \ \#$$

$$\text{Info}_{\text{student}}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times \text{Info}(D_j)$$

$$= \frac{7}{14} I(3,4) + \frac{7}{14} I(6,1)$$

$$= \frac{7}{14}\left[-\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right] + \frac{7}{14}\left[-\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right]$$

$$= \frac{7}{14}(0.524 + 0.461) + \frac{7}{14}(0.191 + 0.401)$$

$$= \frac{7}{14}(0.985) + \frac{7}{14}(0.592)$$

$$= 0.493 + 0.296$$

$$= 0.789 \quad \#$$

$$\text{Info}_{\text{credit}}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times \text{Info}(D_j)$$

$$= \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3)$$

$$= \frac{8}{14}\left[-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right] + \frac{6}{14}\left[-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right]$$

$$= \frac{8}{14}(0.311 + 0.5) + \frac{6}{14}(0.5 + 0.5)$$

$$= \frac{8}{14}(0.811) + \frac{6}{14}$$

$$= 0.464 + 0.429$$

$$= 0.893 \quad \#$$

Gain

Gain (age) $= \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.940 - 0.649 = 0.291$

$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.940 - 0.912 = 0.028$

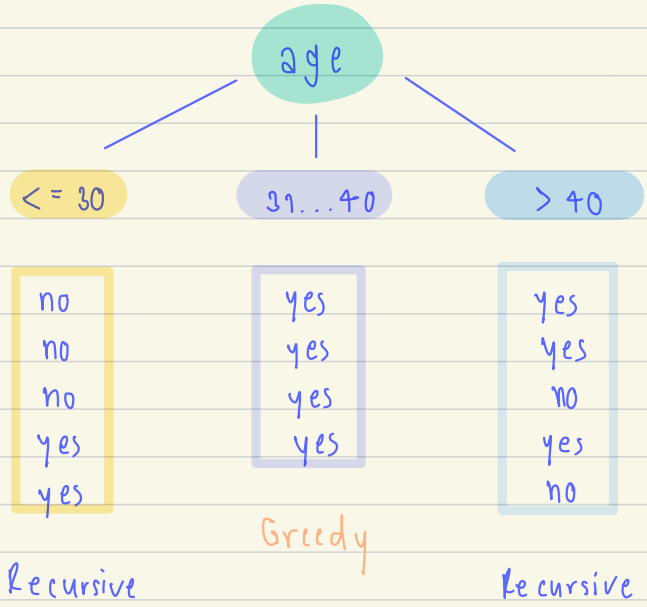
$\text{Gain (Student)} = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.940 - 0.789 = 0.151$

$\text{Gain (credit\_rating)} = \text{Info}(D) - \text{Info}_{\text{credit}}(D) = 0.940 - 0.893 = 0.047$

∴ เลือก **Gain (age)** เพราะมีค่าเยอะที่สุด แปลว่าเป็นทางเลือกที่ดีที่สุด

Training data set: Who buys computer?

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



F_1 age <= 30

age	income	student	credit	buys
<= 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

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

$$= I(2,3)$$

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

$$= 0.971$$

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

$$= \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] + \frac{1}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right]$$

หาค่าไม่ได้ หาค่าไม่ได้

$$= 0.4 \quad \#$$

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

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

$$= 0 \quad \#$$

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

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

$$= 0.551 + 0.4$$

$$= 0.951 \quad \#$$

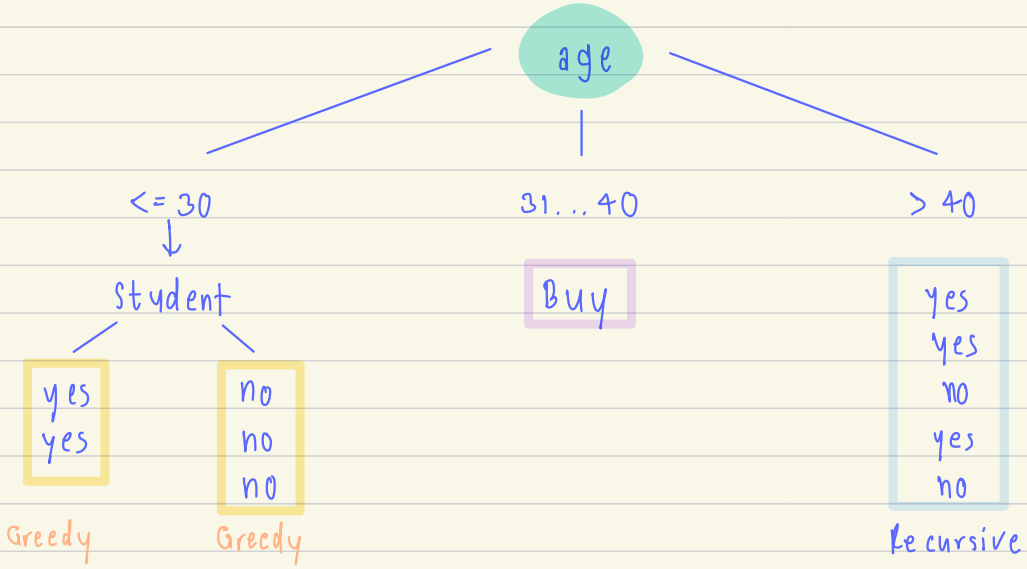
Gain

$Gain(income) = Info(D) - Info_{income}(D) = 0.971 - 0.4 = 0.571$

$Gain(student) = Info(D) - Info_{student}(D) = 0.971 - 0 = 0.971$

$Gain(credit) = Info(D) - Info_{credit}(D) = 0.971 - 0.951 = 0.020$

∴ Gain ที่มากที่สุด คือ Gain(student)



F_2 age > 40

age	income	student	credit	buys
>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

$$Info(D) = I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

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

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

$$= 0.551 + 0.4$$

$$= 0.951 \quad \#$$

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

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

$$= 0.4 + 0.551$$

$$= 0.951 \quad \#$$

$$Info_{credit}(D) = \frac{3}{5}I(3,0) + \frac{2}{5}I(1,1)$$

หาค่าไม่ได้

$$= \frac{3}{5}\left[-\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.4 \quad \#$$

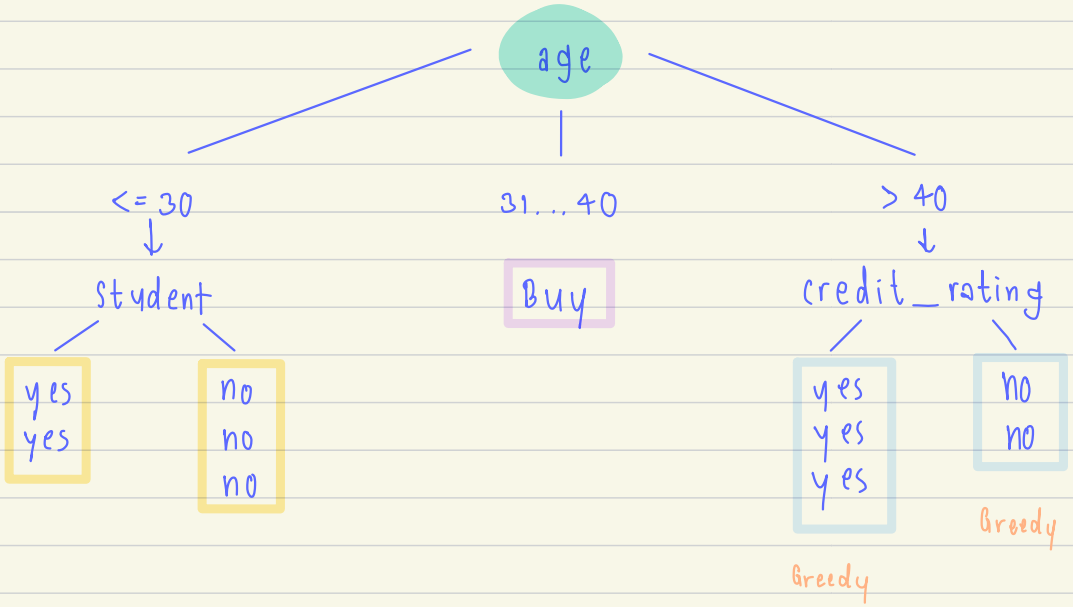
Gain

Gain(income) = Info(D) - $Info_{income}(D)$ = 0.971 - 0.951 = 0.2

Gain(student) = Info(D) - $Info_{student}(D)$ = 0.971 - 0.951 = 0.2

Gain(credit) = Info(D) - $Info_{credit}(D)$ = 0.971 - 0.4 = 0.571

∴ Gain ที่มากที่สุด คือ Gain(credit)



Decision Tree Induction