· Class

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

- $I(9,5)$

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

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

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

$= 0.940$

· Feature

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

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

                         0   บริสุทธิ์

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

$= \dfrac{5}{14}(0.529 + 0.442 + \dfrac{4}{14}(0) + \dfrac{5}{14}(0.442 + 0.529)$

$= \dfrac{5}{14}(0.971) + ไม่เอาไม่ได้ + \dfrac{5}{14}(0.971)$

$= 0.347 + 0.347 = 0.694$

$Gain(age) = Info(D) - Info_{age}(D)$

$= 0.940 - 0.649 = 0.291$

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

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

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

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

$= 0.231 + 0.394 + 0.231 = 0.912$

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

$= 0.940 - 0.912 = 0.028$

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

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

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

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

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

$= 0.493 + 0.296 = 0.789$

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

$= 0.940 - 0.893 = 0.047$

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

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

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

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

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

$= 0.464 + 0.429 = 0.893$

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

$= 0.940 - 0.893 = 0.047$

∴ เลือก $Gain(age)$ เพราะมีค่ามากที่สุด

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

age

$< \cdot 30$	$31...40$	>40
no	yes	yes
no	yes	yes
no	yes	no
yes	yes	yes
yes		no
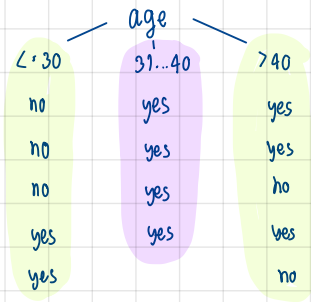
| Recursive | Greedy | Recursive |

F_1 age $< \cdot 30$

age	income	student	credit	buys
$< \cdot 30$	high	no	fair	no
$< \cdot 30$	high	no	excellent	no
$< \cdot 30$	medium	no	fair	no
$< \cdot 30$	low	yes	fair	yes
$< \cdot 30$	medium	yes	excellent	yes

$$\text{Info}(D) = \sum_{i=1}^{\aleph} p_i \log_2 (p_i)$$
$$= I(9,3)$$
$$= -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \log_2 \left(\frac{3}{5}\right) = 0.971$$

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

$$= 0.4$$

$$\text{Gain}_{income} = \text{Info}(D) - \text{Info}_{income}(D)$$
$$= 0.971 - 0.4 = 0.571$$

$$\text{Info}_{student}(D) = \frac{3}{5} \div (0,3) + \frac{2}{5} I(2,0)$$
$$= \frac{3}{5} \left[-\frac{0}{3} \log_2 \left(\frac{0}{3}\right) - \frac{3}{3} \log_2 \left(\frac{3}{3}\right)\right] + \frac{2}{5} \left[-\frac{2}{2} \log_2 \left(\frac{2}{2}\right) - \frac{0}{2} \log_2 \left(\frac{0}{2}\right)\right]$$

$$= 0$$

$$\text{Gain}_{student} = \text{Info}(D) - \text{Info}_{student}(D)$$
$$= 0.971 - 0 = 0.971$$

$$\text{Info}_{credit}(D) = \frac{3}{5} I(1,2) + \frac{2}{5} I(1,1)$$
$$= \frac{3}{5} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right)\right] + \frac{2}{5} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right)\right]$$

$$= 0.551 + 0.4 = 0.951$$

$$\text{Gain}_{credit} = \text{Info}(D) - \text{Info}_{credit}(D)$$
$$= 0.971 - 0.951 = 0.020$$

∴ Gain มากที่สุด คือ Gain $student$

age
            ┌─────────────────────────┼─────────────────────┐
         ≤30   31...40   >40
           │   Buy   yes
        student   yes
      yes   no   no
      yes   no   yes
              no   no
    Greedy   Greedy
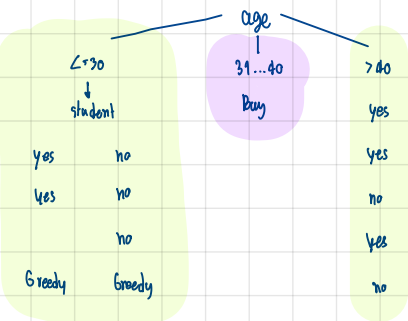
                          Recursive

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

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

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

$\quad = 0.971$

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

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

$\quad = 0.551 + 0.4 = 0.951$

$Gain_{income} = Info(D) - Info_{income}(D)$

$\quad = 0.971 - 0.951 = 0.02$ ✲

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

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

$\quad = 0.4 + 0.551 = 0.951$

$Gain_{student} = Info(D) - Info_{student}(D)$

$\quad = 0.971 - 0.951 = 0.02$ ✲

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

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

$\quad = 0.4$

$Gain_{credit} = Info(D) - Info_{credit}(D)$

$\quad = 0.971 - 0.4 = 0.571$ ✲

                                    age
            ┌─────────────────────────┼─────────────────────┐
         ≤30   31...40   >40
           │   Buy   credit_rating
        student   ┌──────┴──────┐
      yes   no   yes   no
      yes   no   yes   no
              no   yes
    Greedy   Greedy   Greedy   Greedy

                Decision   Tree   Induction
                                    age
            ┌─────────────────────────┼─────────────────────┐
         ≤30   31...40   >40
           │   Buy   credit_rating
        student   ┌──────┴──────┐
      ┌──────┴──────┐   no   yes
     no   yes   not-Buy
  Not-Buy   Buy