สุนิสา อุดมพันธุ์ 6430212 78-8

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

- Expected information (entropy) needed to classify a tuple in D:

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

- Information needed (after using A to split D into v partitions) to classify D:

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

- Information gained by branching on attribute A

$$Gain(A) = Info(D) - Info_A(D)$$

- คำนวณ Class

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

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

$$= 0.41 + 0.53$$

$$Info(D) = 0.940 \ \#$$

- คำนวณ Feature $Info_A(D)$

สูตร $$Info_A(D) = \sum_{i=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

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

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

$$\frac{5}{14}\left[-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right)\right]$$

$$= 0.34676 + 0.34676$$

$$Info_{age}(D) = 0.694 \ \#$$

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

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

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

$$= 0.2857 + 0.3935 + 0.2317$$

$$Info_{income}(D) = 0.911 \ \#$$

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

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

$= \frac{6}{14}\left[0.5 + 0.5\right] + \frac{8}{14}\left[0.3113 + 0.5\right]$

$= \left(\frac{6}{14} \times 1\right) + \left(\frac{8}{14} \times 0.8713\right)$

$= 0.4285 + 0.4636$

$Info_{credit}(D) = 0.892$

□ $Info_{stud.}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4)$

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

$= \frac{7}{14}\left[0.191 + 0.401\right) + \frac{7}{14}\left[0.524 + 0.461\right]$

$= 0.296 + 0.493$

$Info_{stu}(D) = 0.788$ #

$Gain(A) = Info(D) - Info_A(D)$

$Gain(age) = 0.940 - 0.694 = 0.246$ ✱ Root node

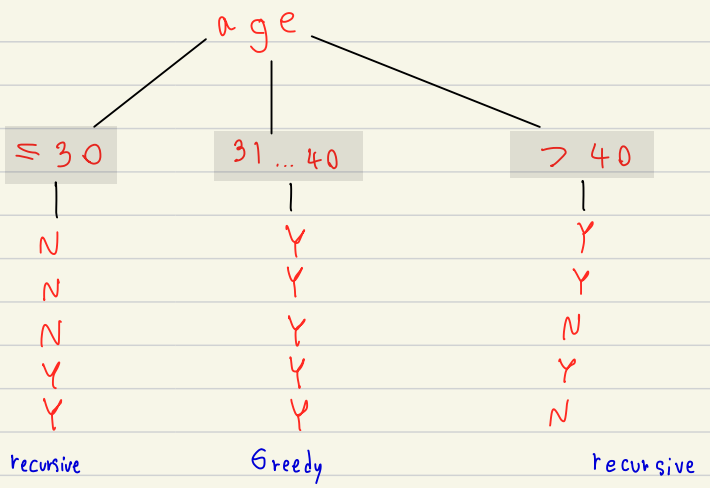
$Gain(income) = 0.940 - 0.911 = 0.029$

$Gain(student) = 0.940 - 0.788 = 0.152$

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

Recureive age ≤ 30

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



- คำนวน Class

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

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

$$= 0.5288 + 0.4422$$

$$Info(D) = 0.9710$$

- คำนวน Feature

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

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

$$= \frac{2}{5}(\text{ขนไม่ใช่}) + \frac{2}{5}(0.5+0.5) + \frac{1}{5}(\text{ขนใช่ไม่ใช่})$$

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

$$Info_{income}(D) = 0.4 \quad \#$$

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

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

$$= \frac{2}{5}[\text{ขนใช่ไม่ใช่}] + \frac{3}{5}[\text{ขนไม่ใช่ไม่ใช่}]$$

$$= 0 + 0$$

$$Income_{stu}(D) = 0 \quad \#$$

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

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

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

$= \left[\frac{2}{5}\times 1\right] + \left[\frac{3}{5}\times 0.918\right]$

$= 0.4 + 0.551$

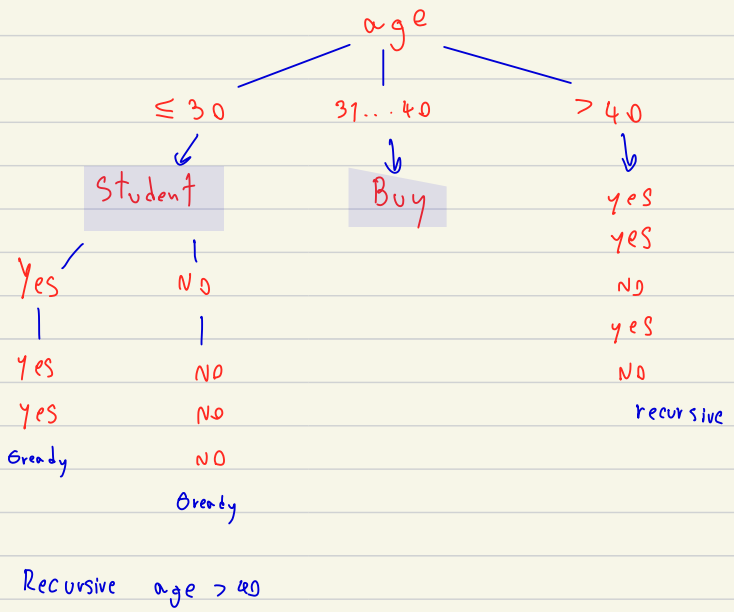
Info credit (D) $= 0.951$ #

Gain (income) $= 0.9710 - 0.4 = 0.571$

$= 0.9710 - 0 = 0.971$ ✅ มากสุด

$= 0.9910 - 0.9501 = 0.020$

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



Recursive age > 40

0 Mium Class

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

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

$= 0.442 + 0.528$

Info (D) $= 0.971$ #

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

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

$$= \frac{3}{5}[0.389 + 0.528] + \frac{2}{5}[0.5 + 0.5]$$

$$= 0.5509 + 0.4$$

$$Info_{income}(D) = 0.9509 \quad \#$$

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

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

$$= \frac{2}{5}[\text{ไม่ได้ใช้}] + \frac{3}{5}[\text{ได้ใช้แน่}]$$

$$Info_{credit}(D) = 0 \quad \#$$

$Gain(income) = 0.710 - 0.9509 = 0.0201$

$Gain(credit\_rating) = 0.9710 - 0 = 0.9710$ ✅ Gain มากสุด

- **Resulting tree:**