ตะวัน เข็มเขียวเพชร
643020501-6

>40	3	2	0.971

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

no = 5/14

yes = 9/14

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

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

$= -[-0.53 + -0.409]$

$= 0.939$

คำนวณ แต่ละ feature

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

age

<= 30

$j = 5$

$Info\ D_j = -\left[\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right] = -(-0.44 + -0.52) = 0.96$

31...40 $\Rightarrow D_j = 4$

$$Info\ D_j = -\left[\frac{4}{4}\log_2\left(\frac{4}{4}\right) + \frac{0}{4}\log_2\left(\frac{0}{4}\right)\right] = \text{undefined}$$

(undefined)

>40 $\Rightarrow D_j = 5$

$$Info\ D_j = -\left[\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right] = 0.96$$

$$Info_{Age}(D) = \left[\frac{5}{14}(0.96) + \frac{4}{14}(\text{undefined}) + \frac{5}{14}(0.96)\right]$$

$Info_{Age}(D) = 0.69$

Income

high medium low

$$Info_{income}(D) = \sum_{j=1} \frac{|D_j|}{|D|} \times info(D_j)$$

$$= \left[\frac{4}{14} \times info(2,2) + \frac{6}{14} \times info(4,2) + \frac{4}{14} \times info(3,1)\right]$$

(high) (medium) (low)

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

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

$$= \frac{4}{14}[0.5 + 0.5] + \frac{6}{14}[0.38 + 0.52] + \frac{4}{14}[0.31 + 0.5]$$

$Info_{income}(D) = 0.91$

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

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

$$= \frac{7}{14}\left[0.19 + 0.40\right] + \frac{7}{14}\left[0.52 + 0.46\right]$$

$$Info_{student}(D) = 0.78$$

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

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

$$= \frac{6}{14}\left[0.5 + 0.5\right] + \frac{8}{14}\left[0.31 + 0.5\right]$$

$$Info_{credit}(D) = 0.89$$

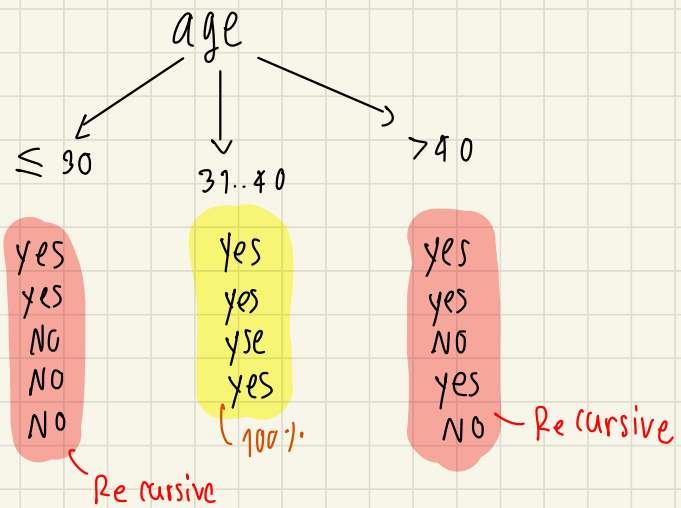
ดัง $Gian(A) = Info(D) - Info_A(D)$

$Gian(age) = 0.94 - 0.69 = 0.246$ → ดัง Gian ที่มากที่สุด เป็น root node

$Gian(income) = 0.94 - 0.91 = 0.029$

$Gian(student) = 0.94 - 0.78 = 0.162$

$Gian(credit) = 0.94 - 0.89 = 0.048$



พิจารณา age ≤ 30

age	income	student	credit_rating	buys_computer
>40	3	2	0.971	
<=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

no = 3
yes = 2

$$Info_{<=30}(D) = I(2,3)$$

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

$$Info(D) = 0.97$$

หา feature ที่เหมาะสม

(high: undefind; Medium; Low: undefind)

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

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

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

$$Info_{income}(D) = 0.4$$

(yes: undefind; NO: undefind)

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

$Info_{stadent}(D)$ = หาค่าไม่ได้ ✓

(ex; fair)

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

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

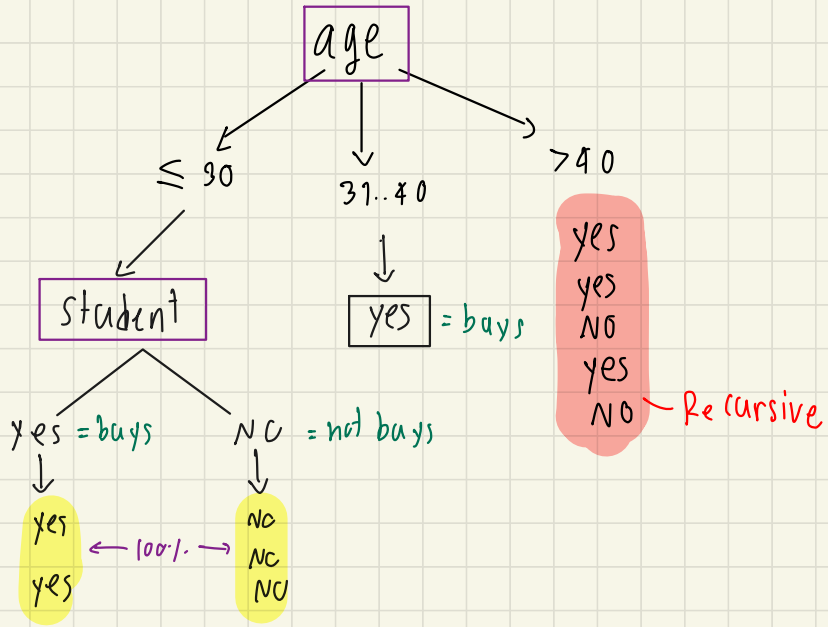
$$= \frac{2}{5}[0.5+0.5] + \frac{3}{5}[0.52 + 0.38]$$

$$= 0.95$$

Gian (income) = Info (D) - $Info_{income}$ (D) = 0.97 - 0.4 = 0.57

Gian (student) = Info (D) - $Info_{student}$ (D) = 0.97

Gian (credit) = Info (D) - $Info_{credit}$ (D) = 0.97 - 0.95 = 0.02



พิจารณา >40

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

$$Info_{>40}(D) = I(3,2)$$

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

$$Info(D) = 0.97$$

คน feature ตัวไหนดีที่สุด

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

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

$$= \frac{3}{5}[0.38 + 0.52] + \frac{2}{5}[0.5 + 0.5]$$

$$= \frac{3}{5}(0.91) + \frac{2}{5}(1)$$

$Info_{income}(D) = 0.95$

---

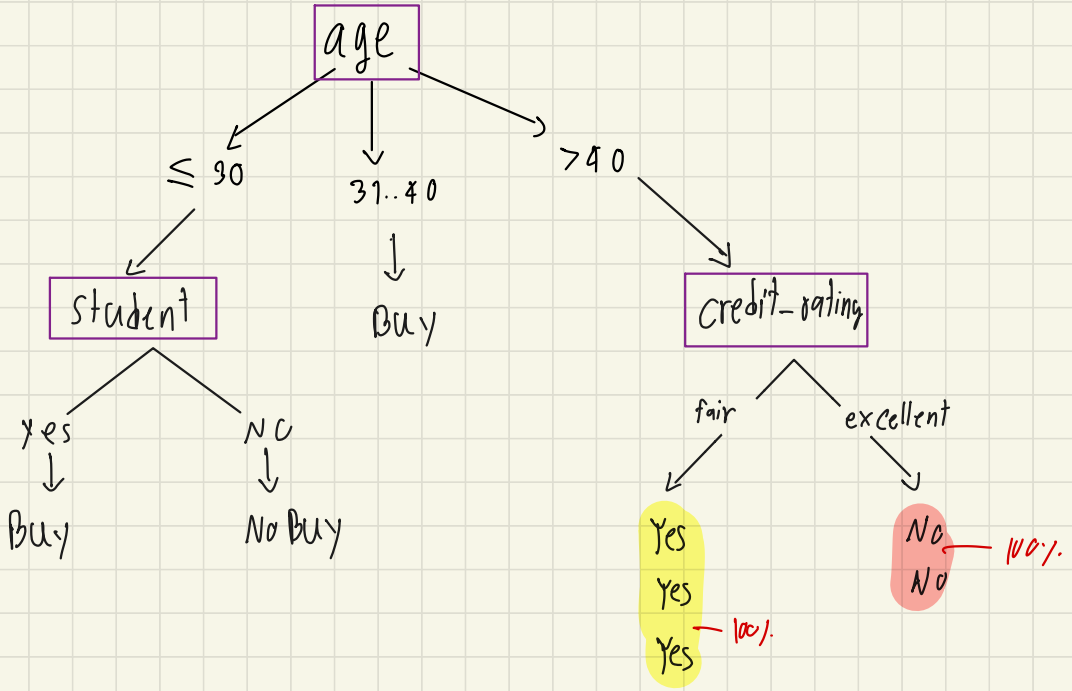
undefind undefind

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

$Info_{credit}(D)$ = คำนวณไม่ได้

$Gain(income) = 0.97 - 0.95 = 0.02$

$Gain(credit\_rating) = 0.97 = 0.97$ → gain ตัวที่สุด



final Decision Tree