นางสาวเพ็ญพิชชา วรรณชุมาตร์ 643020063-4

HW2 : Decision Tree

age	income	student	Credit_rat
≤ 30 (2,3) 5	high (2,2) 4	yes (6,1) 7	excellent (3,3) 6
31...40 (4,0) 4	medium (4,2) 6	no (3,4) 7	fair (6,2) 8
>40 (3,2) 5	low (3,1) 4		

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

x : features — Class

Class p คือ buys_computer = yes
Class n คือ buys_computer = no

1) คำนวณ Class

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

คำนวณ $= I(9,5)$

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

$= 0.41 + 0.53$

∴ $Info(D) = 0.940$ ✱

2) คำนวณ feature

2.1) จากสูตร $Info_A(D) = \sum_{j=1}^{y}\left[\frac{|D_j|}{|D|} \times Info(D_j)\right]$

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

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

$= \frac{5}{14}\left[0.5287 + 0.44217\right] + \left[\frac{4}{14} \times \text{ไม่ต้องคำนวณ}\right] + \frac{5}{14}\left[0.44217 + 0.52877\right]$

$= \left[\frac{5}{14} \times 0.9709\right] + \left[\frac{5}{14} \times 0.9709\right]$

$= 0.34676 + 0.34676$

$= 0.69352 \approx 0.694$

∴ $Info_{age}(D) = 0.694$ ✱

2.2) จากสูตร $Info_A(D) = \sum_{j=1}^{y} \left[\frac{|D_j|}{|D|} \times Info(D_j) \right]$

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

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

$$= \frac{4}{14}\left[0.5 + 0.5 \right] + \frac{6}{14}\left[0.3899 + 0.5283 \right] + \frac{4}{14}\left[0.3112 + 0.5 \right]$$

$$= \left[\frac{4}{14} \times 1 \right] + \left[\frac{6}{14} \times 0.9182 \right] \quad \left[\frac{4}{14} \times 0.8112 \right]$$

$$= 0.2857 + 0.3935 + 0.2317$$

$$= 0.9109 \approx 0.911$$

∴ $Info_{income}(D) = 0.911$ #

2.3) จากสูตร $Info_A(D) = \sum_{j=1}^{y} \left[\frac{|D_j|}{|D|} \times Info(D_j) \right]$

$$Info_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,5)$$

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

$$= \frac{7}{14}\left[0.1906 + 0.4010 \right] + \frac{7}{14}\left[0.5238 + 0.4613 \right]$$

$$= \left[\frac{7}{14} \times 0.5916 \right] + \left[\frac{7}{14} \times 0.9851 \right]$$

$$= 0.2958 + 0.4925$$

$$= 0.78835 \approx 0.788$$

∴ $Info_{student}(D) = 0.788$ #

2.4)

$$Info_A(D) = \sum_{j=1}^{y} \left[\frac{|D_j|}{|D|} \times Info(D_j) \right]$$

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

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

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

$$= \left[\frac{6}{14} \times 1 \right] + \left[\frac{8}{14} \times 0.8113 \right]$$

$$= 0.4285 + 0.4636$$

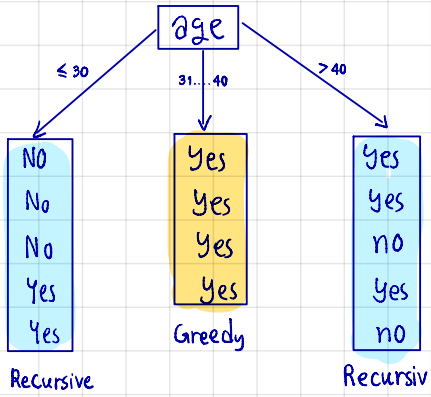
$$= 0.8921 \approx 0.892$$

∴ $Info_{credit}(D) = 0.892$ #

Gain(A)	$Info(D) - Info_A(D)$	Gain
age	0.940 - 0.694	0.246
income	0.940 - 0.911	0.029
student	0.940 - 0.788	0.152
Credit_rating	0.940 - 0.892	0.048

→ เลือก Gain ที่มากที่สุดเป็น root node

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



ที่ Recursive age ≤ 30

age	income	student	Credit
≤ 30 (2,3) 5	high (0,2) 2	yes (2,0) 2	excellent (1,1) 2
	medium (1,1) 2	no (0,3) 3	fair (2,1) 3
	low (1,0) 1		

1) คำนวน Class

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

$$= 0.5288 + 0.4422$$

$$\therefore Info(D) = 0.9710 \ \text{※}$$

2) คำนวน feature

จากสูตร $Info_A(D) = \sum_{j=1}^{y}\left[\frac{|D_j|}{|D|} \times Info(D_j)\right]$

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

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

$$= \left[\frac{2}{5} \times \text{ทศนิยมไม่ได้}\right] + \frac{2}{5}[0.5+0.5] + \left[\frac{1}{5} \times \text{ทศนิยมไม่ได้}\right]$$

$$\therefore Info_{income}(D) = 0.4 \ \text{※}$$

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

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

$$= \left[\frac{2}{5} \times \text{ทศนิยมไม่ได้}\right] + \left[\frac{3}{5} \times \text{ทศนิยมไม่ได้}\right]$$

$$\therefore Info_{student}(D) = 0 \ \text{※}$$

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

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

$= \frac{2}{5}[0.5+0.5] + \frac{3}{5}[0.5283 + 0.3899]$

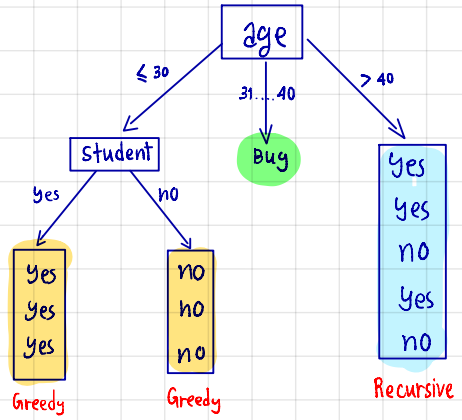
$= \left[\frac{2}{5} \times 1\right] + \left[\frac{3}{5} \times 0.9182\right]$

$= 0.4 + 0.5509$

∴ $Info_{credit}(D) = 0.9509$ ※

Gain(A)	$Info(D) - Info_A(D)$	Gain
income	0.9710 − 0.4	0.5710
student	0.9710 − 0	0.9710 → เลือก Gain ที่มากที่สุด
Credit_rating	0.9710 − 0.9509	0.0201

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



ที่ Recursive age > 40

1) คำนวณ Class

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

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

$= 0.4422 + 0.5288$

∴ $Info(D) = 0.971$ ※

age	income	student	Credit
	high (0,0)	yes (2,1)	excellent (0,2)
>40	0	3	2
(3,2)	medium (2,1)	no (1,1)	fair (3,0)
5	3	2	3
	low (1,1)		
	2		

2) คำนวณ feature

จากสูตร $Info_A(D) = \sum_{j=1}^{y} \left[\frac{|D_j|}{|D|} \times Info(D_j) \right]$

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

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

$= \frac{3}{5}\left[0.3899 + 0.5283 \right] + \frac{2}{5}\left[0.5 + 0.5 \right]$

$= \left[\frac{3}{5} \times 0.9182 \right] + \left[\frac{2}{5} \times 1 \right]$

$= 0.5509 + 0.4$

∴ $Info_{income}(D) = 0.9509$ #

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

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

$= \frac{3}{5}\left[0.3899 + 0.5283 \right] + \frac{2}{5}\left[0.5 + 0.5 \right]$

$= \left[\frac{3}{5} \times 0.9182 \right] + \left[\frac{2}{5} \times 1 \right]$

$= 0.5509 + 0.4$

∴ $Info_{student}(D) = 0.9509$ #

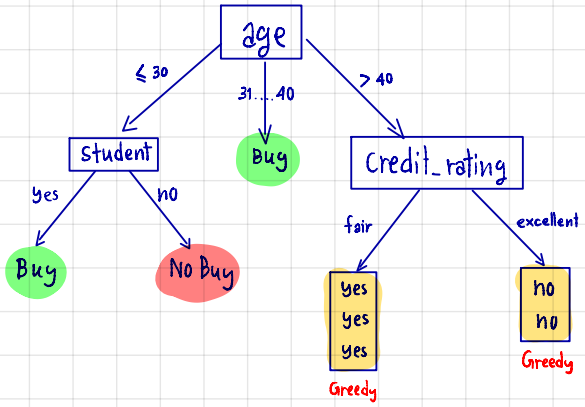
2.3) $Info_{credit}(D) = \frac{2}{5} I(0,2) + \frac{3}{5} I(3,0)$

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

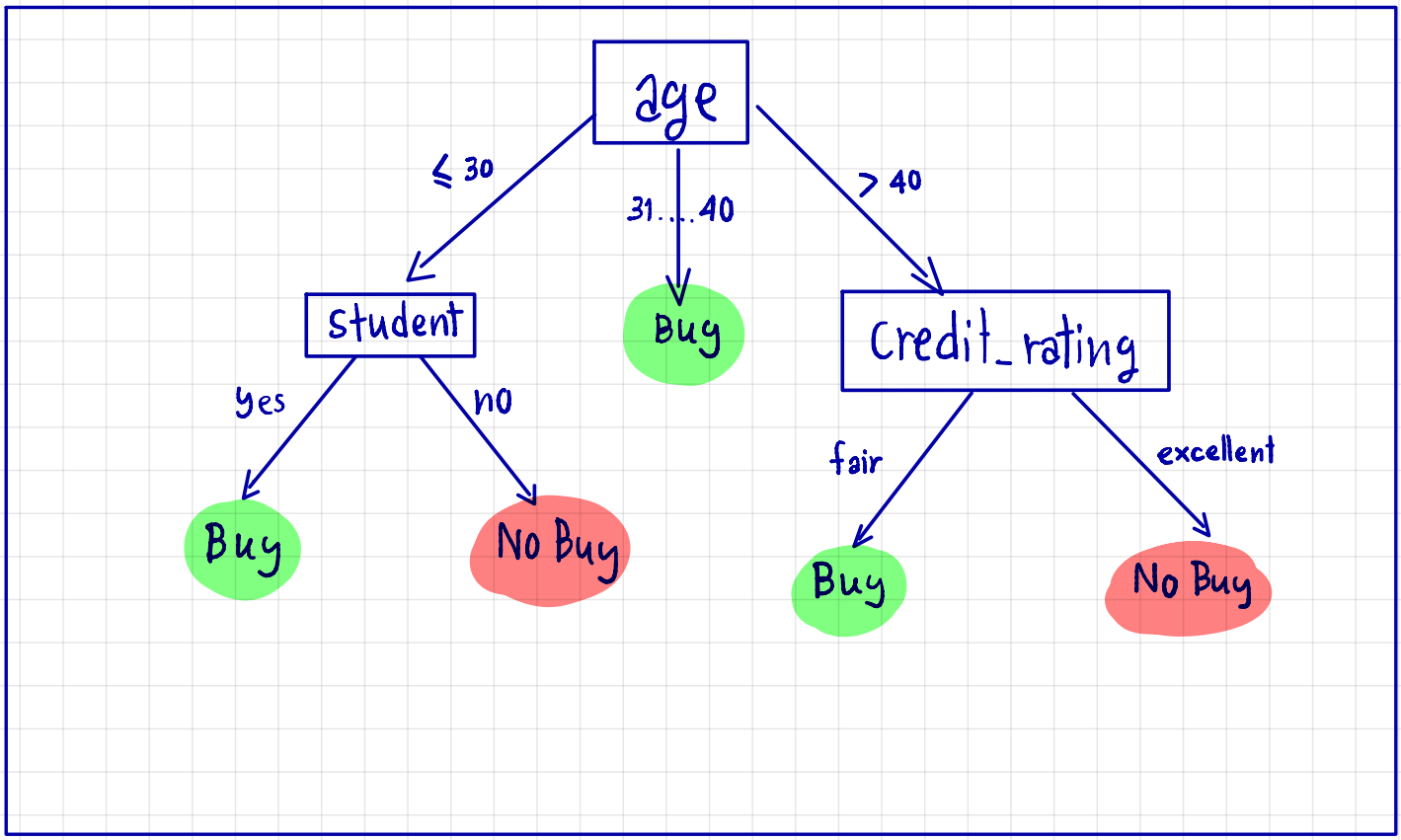
$= \left[\frac{2}{5} \times \text{ทศนิยมไม่ได้} \right] + \left[\frac{3}{5} \times \text{ทศนิยมไม่ได้} \right]$

∴ $Info_{credit}(D) = 0$ #

Gain(A)	$Info(D) - Info_A(D)$	Gain
income	0.971 - 0.9509	0.0201
student	0.971 - 0.9509	0.0201
Credit_rating	0.971 - 0	0.971 → เลือก Gain ที่มากที่สุด


age
≤ 30
31....40
> 40
Student
Buy
Credit_rating
yes
no
fair
excellent
Buy
No Buy
yes
yes
yes
Greedy
no
no
Greedy

final


age
≤ 30
31....40
> 40
Student
Buy
Credit_rating
yes
no
fair
excellent
Buy
No Buy
Buy
No Buy