นางสาวณัธกมล อัมฤตานนท์
6330204394

Class (buys)

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

= 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

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

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

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

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

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

= 0.347 + 0.347

= 0.694

Gain (age) = Info (D) − $Info_{age}(D)$ = 0.940 − 0.649 = 0.291

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

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

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

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

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

= 0.286 + 0.39 + 0.232 = 0.912

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

Info student (D)

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

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

$$= \frac{7}{14}\left[\frac{-3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right] + \frac{7}{17}\left[\frac{-6}{7}\log\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right]$$

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

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

$$= 0.493 + 0.296 = 0.789$$

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

Info credit (D)

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

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

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

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

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

$$= 0.464 + 0.429$$

$$= 0.893$$

Gain (credit_rating) $= \text{Info}(D) - \text{Info}_{\text{credit}}(D)$

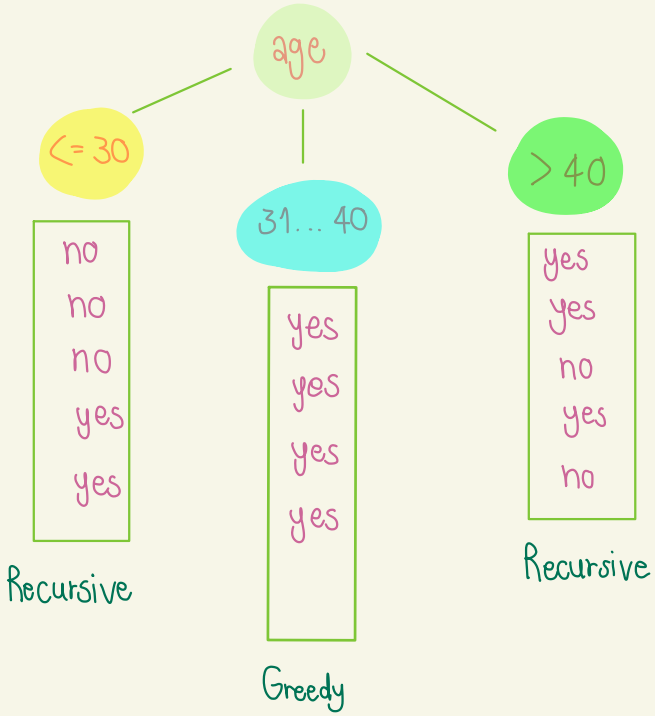
$$= 0.94 - 0.893$$

$$= 0.047$$

ดังนั้นเราจะเลือก Gain (age) เพราะว่าค่าที่เยอะที่สุดซึ่งแปลว่าเป็นทางเลือกที่ดีที่สุด

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

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

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

$$= 0.4$$

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

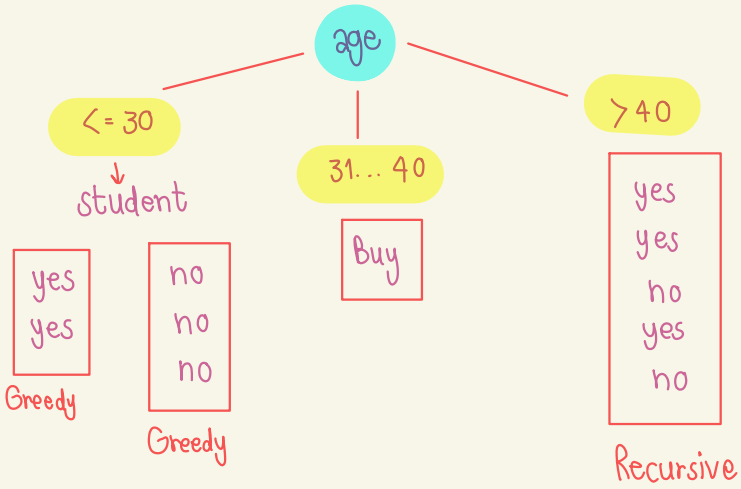
$$\text{Info}_{\text{student}}(D) = \frac{3}{5} \pm (0,3) + \frac{2}{5} \pm (2,0)$$
$$= \frac{3}{5}\left[-\frac{0}{3}\log_2\left(\frac{0}{3}\right) - \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right] + \frac{2}{5}\left[-\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right]$$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

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

$$\text{Info}_{\text{credit}}(D) = \frac{3}{5} \pm (1,2) + \frac{2}{5} \pm (1,1)$$
$$= \frac{3}{5}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right)\right]$$
$$= 0.551 + 0.4 = 0.951$$

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

ดังนั้น Gain ที่มากที่สุด คือ Gain (student)



F_2 age > 40

age	income	student	credit	buy
>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

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

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

$$= 0.551 + 0.4 = 0.951$$

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

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

$$= 0.4 + 0.551 = 0.951$$

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

$$\text{Info}_{\text{credit}}(D) = \frac{3}{5} \pm (3,0) + \frac{2}{5} \pm (1,1)$$

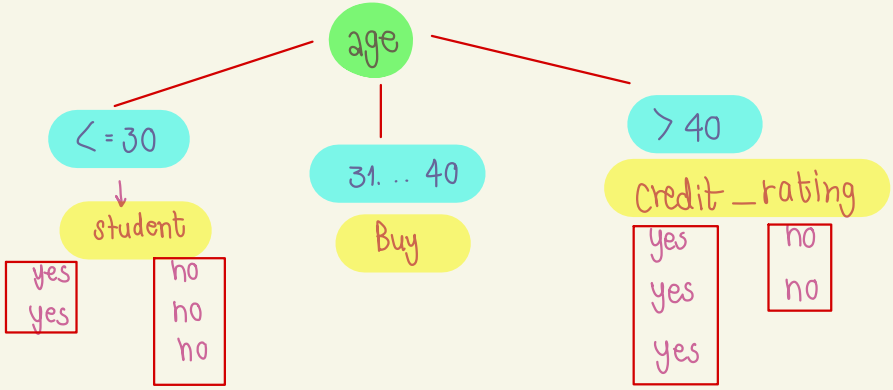
$$= \frac{3}{5}\left[-\frac{3}{5}\log_2\left(\frac{2}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.4$$

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

$$= 0.571$$

ดังนั้น Gain ที่มีมากที่สุด คือ Gain (credit)



Decision Tree Induction