

HW 2 → Attribute Selection with Information Gain & Decision Tree Induction

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	p_i	n_i	$I(p_i, n_i)$	$p_i + n_i$
<=30	2	3	0.971	5
31...40	4	0	0	4
>40	3	2	0.971	5

$$I(2,3) = -\frac{2}{5} \log_2(\frac{2}{5}) - \frac{3}{5} \log_2(\frac{3}{5}) = 0.971$$

$$I(4,0) = -\frac{4}{4} \log_2(\frac{4}{4}) - \frac{0}{4} \log_2(\frac{0}{4}) = 0$$

$$I(3,2) = -\frac{3}{5} \log_2(\frac{3}{5}) - \frac{2}{5} \log_2(\frac{2}{5}) = 0.971$$

income	p_i	n_i	$p_i + n_i$	$I(p_i, n_i)$
high	2	2	4	1
medium	4	2	6	0.918
low	3	1	4	0.811

↳ Class P: buys_computer = "yes"

↳ Class N: buys_computer = "no"

$$\text{Info}(D) = I(9,5) = -\frac{9}{14} \log_2(\frac{9}{14}) - \frac{5}{14} \log_2(\frac{5}{14}) = 0.940$$

$$\begin{aligned} \text{Info}_{\text{age}}(D) &= \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) \\ &= \frac{5}{14} (0.971) + \frac{4}{14} (0) + \frac{5}{14} (0.971) \\ &= 0.694 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) \\ &= \frac{4}{14} (1) + \frac{6}{14} (0.918) + \frac{4}{14} (0.811) \\ &= 0.911 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{student}}(D) &= \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) = \frac{7}{14} (0.592) + \frac{7}{14} (0.985) \\ &= 0.789 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{credit_rating}}(D) &= \frac{6}{14} I(3,3) + \frac{8}{14} I(6,2) = \frac{6}{14} (1) + \frac{8}{14} (0.811) \\ &= 0.892 \end{aligned}$$

$$I(2,2) = -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) = 1$$

$$I(4,2) = -\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) = 0.918$$

$$I(3,1) = -\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) = 0.811$$

Student	P_i	n_i	P_i+n_i	$I(P_i, n_i)$
yes	6	1	7	0.592
no	3	4	7	0.981

$$I(6,1) = -\frac{6}{7} \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \log_2\left(\frac{1}{7}\right) = 0.592$$

$$I(3,4) = -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) = 0.985$$

Credit_rating	P_i	n_i	P_i+n_i	$I(P_i, n_i)$
excellent	3	3	6	1
fair	6	2	8	0.811

$$I(3,3) = -\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) = 1$$

$$I(6,2) = -\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) = 0.811$$

$$\begin{aligned} \text{Gain(age)} &= \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.940 - 0.694 \\ &= 0.246 \times \text{root node} \end{aligned}$$

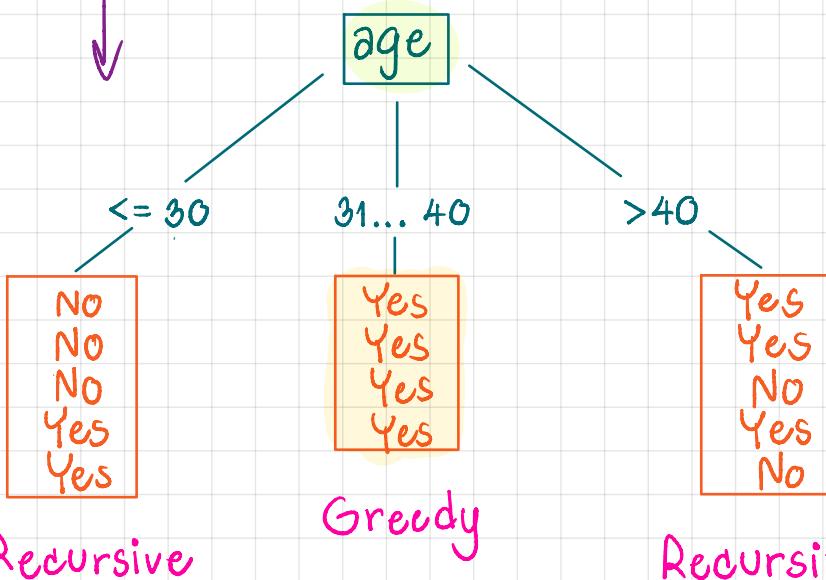
$$\begin{aligned} \text{Gain(income)} &= \text{Info}(D) - \text{Info}_{\text{income}}(D) \\ &= 0.940 - 0.911 = 0.029 \end{aligned}$$

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

$$\begin{aligned} \text{Gain(credit_rating)} &= \text{Info}(D) - \text{Info}_{\text{credit_rating}}(D) \\ &= 0.940 - 0.892 = 0.048 \end{aligned}$$

∴ Root Node ဆုံးပါ၏ Gain များစွာ

အကြောင်း၏ Gain(age) များကို 0.246 ×



กรณี age: ≤ 30

$$\text{Info}_{\text{age}: \leq 30}(D) = I(2, 3) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) = 0.971$$

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \frac{2}{5} I(0, 2) + \frac{1}{5} I(1, 1) + \frac{1}{5} I(1, 0) \\ &= \frac{2}{5}(0) + \frac{1}{5}(1) + \frac{1}{5}(0) \\ &= 0.4 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{student}}(D) &= \frac{2}{5} I(2, 0) + \frac{3}{5} I(0, 3) \\ &= \frac{2}{5}(0) + \frac{3}{5}(0) = 0 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{credit_rating}}(D) &= \frac{2}{5} I(1, 1) + \frac{3}{5} I(1, 2) = \frac{2}{5}(1) + \frac{3}{5}(0.918) \\ &= 0.951 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{income}) &= \text{Info}_{\text{age}: \leq 30} - \text{Info}_{\text{income}}(D) \\ &= 0.971 - 0.4 = 0.571 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{student}) &= \text{Info}_{\text{age}: \leq 30} - \text{Info}_{\text{student}}(D) \\ &= 0.971 - 0 = 0.971 \times \text{Decision Node} \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{credit_rating}) &= \text{Info}_{\text{age}: \leq 30} - \text{Info}_{\text{credit_rating}}(D) \\ &= 0.971 - 0.951 = 0.020 \end{aligned}$$

\therefore Decision Node 1 คือ Student เมื่อ Gain มากที่สุด โดยมีค่า Gain เท่ากับ 0.971 \times

income	P_i	n_i	P_i+n_i	$I(P_i+n_i)$
high	0	2	2	0
medium	1	1	2	1
low	1	0	1	0

$$I(0, 2) = -\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) = 0$$

$$I(1, 1) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$I(1, 0) = -\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) = 0$$

Student	P_i	n_i	P_i+n_i	$I(P_i, n_i)$
yes	2	0	2	0
no	0	3	3	0

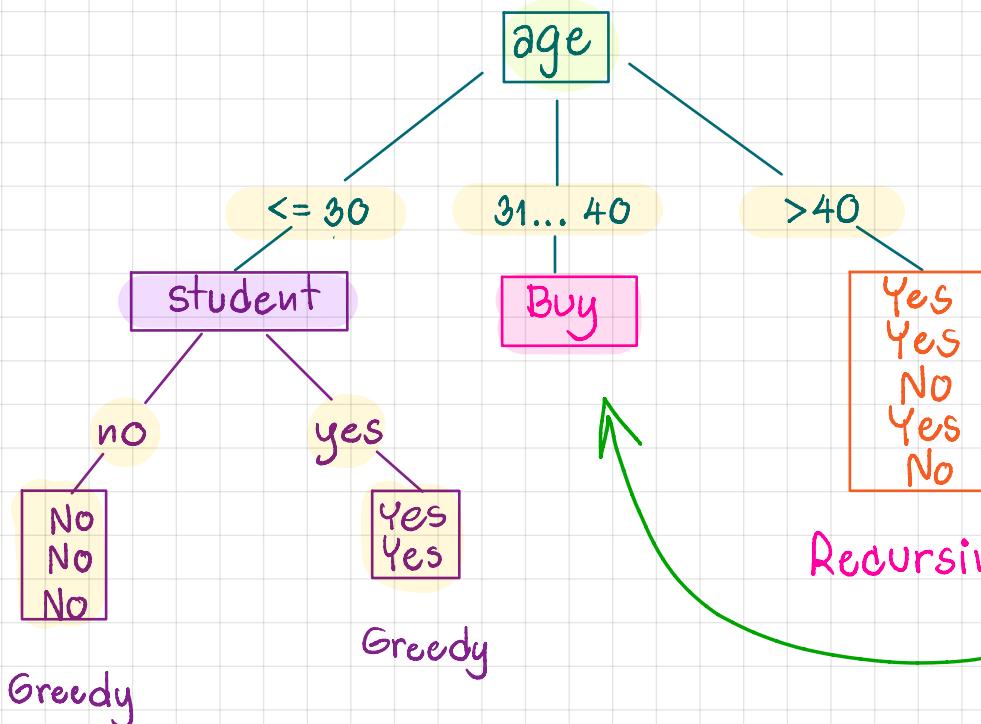
$$I(2, 0) = -\frac{2}{2} \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \log_2\left(\frac{0}{2}\right) = 0$$

$$I(0, 3) = -\frac{0}{3} \log_2\left(\frac{0}{3}\right) - \frac{3}{3} \log_2\left(\frac{3}{3}\right) = 0$$

credit_rating	P_i	n_i	P_i+n_i	$I(P_i, n_i)$
excellent	1	1	2	1
fair	1	2	3	0.918

$$I(1, 1) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$I(1, 2) = -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) = 0.918$$



পর্যাপ্ত age: 31... 40

Infoage: 31... 40 = I(4,0)

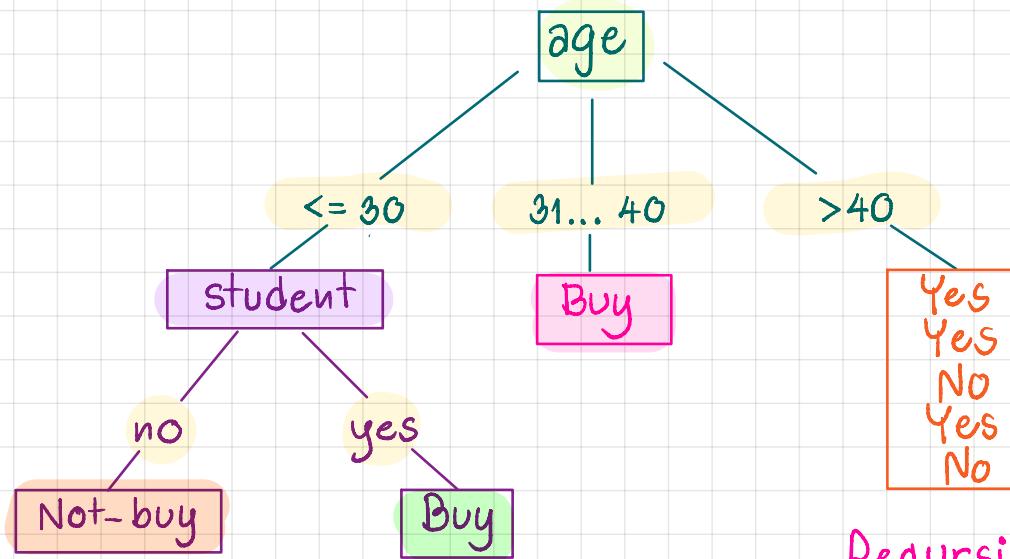
$$= -\frac{4}{4} \log_2\left(\frac{4}{4}\right) - \frac{0}{4} \log_2\left(\frac{0}{4}\right)$$

$$= 0$$

অন্তিম নোট

∴ এই গুরুত্বের age: 31... 40 এর Decision Node

বের পদ্ধতি age: 31... 40 এর yes রেখা Buy
নির্ণয় করো



Recursive

पता : age : > 40

$$\text{Info}_{\text{age}:>40}(D) = I(3,2) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) = 0.971$$

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1) = \frac{3}{5}(0.918) + \frac{2}{5}(1) \\ &= 0.951 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{student}}(D) &= \frac{3}{5}I(1,1) + \frac{2}{5}I(1,1) = \frac{3}{5}(0.918) + \frac{2}{5}(1,1) \\ &= 0.951 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{credit_rating}}(D) &= \frac{2}{5}I(0,2) + \frac{3}{5}I(3,0) = \frac{2}{5}(0) + \frac{3}{5}(0) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{income}) &= \text{Info}_{\text{age}:>40} - \text{Info}_{\text{income}}(D) \\ &= 0.971 - 0.951 = 0.020 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{student}) &= \text{Info}_{\text{age}:>40} - \text{Info}_{\text{student}}(D) \\ &= 0.971 - 0.951 = 0.020 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{credit_rating}) &= \text{Info}_{\text{age}:>40} - \text{Info}_{\text{credit_rating}}(D) \\ &= 0.971 + 0 = 0.971 \quad \times \end{aligned}$$

∴ Decision Node 2 ជំនួយ Credit_rating បានរាយដែល Gain ខាងក្រោម

ទូទៅ Gain នៅក្នុង 0.971 \times

income	P_i	n_i	P_i+n_i	$I(P_i+n_i)$
high	0	0	0	0
medium	2	1	3	0.918
low	1	1	2	1

$$I(0,0) = 0$$

$$I(2,1) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$$

$$I(1,1) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

Student	P_i	n_i	P_i+n_i	$I(P_i, n_i)$
yes	2	1	3	0.918
no	1	1	2	1

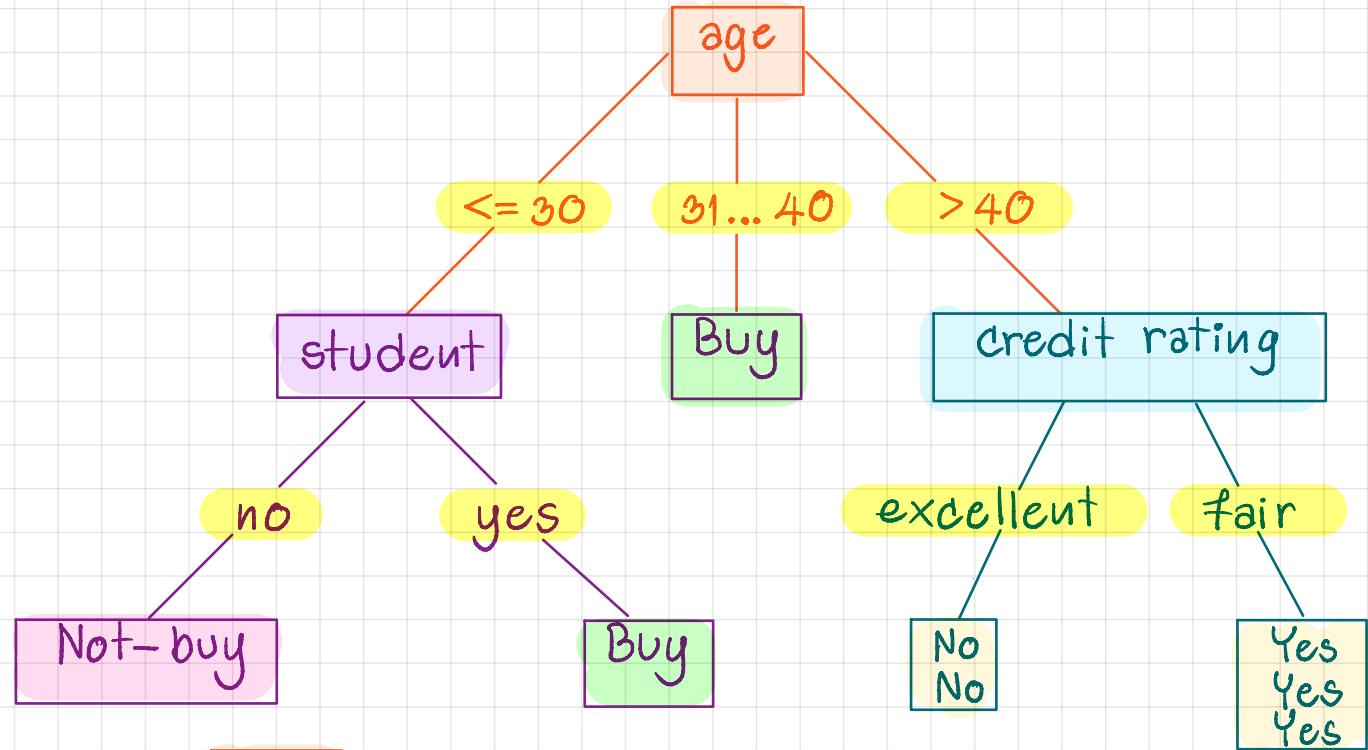
$$I(2,1) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$$

$$I(1,1) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

credit_rating	P_i	n_i	P_i+n_i	$I(P_i, n_i)$
excellent	0	2	2	0
fair	3	0	3	0

$$I(0,2) = -\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) = 0$$

$$I(3,0) = -\frac{3}{3} \log_2\left(\frac{3}{3}\right) - \frac{0}{3} \log_2\left(\frac{0}{3}\right) = 0$$



Resulting tree :

