

ตัวอย่างนี้ → 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

data set 12 of 5

$$\begin{aligned} \text{Gain}(\text{age}) &= \text{Info}(D) - \text{Info}_{\text{age}}(D) \\ &= 0.9183 - 0.7379 \\ &= 0.1804 \end{aligned}$$

$$\text{Info}(D) = I(8,4) = -\frac{8}{12} \log_2\left(\frac{8}{12}\right) - \frac{4}{12} \log_2\left(\frac{4}{12}\right) = 0.9183$$

age	$p_i$	$n_i$	$I(p_i, n_i)$
<=30	2	2	1
31...40	3	0	0
>40	3	2	0.9710

$$\text{Info}_{\leq 30}(D) = I(2,2) = -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) = 1$$

$$\text{Info}_{31...40}(D) = I(3,0) = -\frac{3}{3} \log_2\left(\frac{3}{3}\right) - \frac{0}{3} \log_2\left(\frac{0}{3}\right) = 0$$

$$\text{Info}_{>40}(D) = I(3,2) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) = 0.9710$$

$$\text{Info}_{\text{age}}(D) = \frac{4}{12} I(2,2) + \frac{3}{12} I(3,0) + \frac{5}{12} I(3,2)$$

$$= \frac{4}{12} (1) + \frac{3}{12} (0) + \frac{5}{12} (0.9710)$$

$$= 0.7379$$

income	$p_i$	$n_i$	$I(p_i, n_i)$
high	2	2	1
medium	4	1	0.7219
low	2	1	0.9183

$$\text{Info}_{\text{high}}(D) = I(2,2) = -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) = 1$$

$$\text{Info}_{\text{medium}}(D) = I(4,1) = -\frac{4}{5} \log_2\left(\frac{4}{5}\right) - \frac{1}{5} \log_2\left(\frac{1}{5}\right) = 0.7219$$

$$\text{Info}_{\text{low}}(D) = I(2,1) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.9183$$

$$\text{Info}_{\text{income}}(D) = \frac{4}{12} I(2,2) + \frac{5}{12} I(4,1) + \frac{3}{12} I(2,1)$$

$$= \frac{4}{12} (1) + \frac{5}{12} (0.7219) + \frac{3}{12} (0.9183)$$

$$= 0.8637$$

$$\begin{aligned} \text{Gain}(\text{income}) &= \text{Info}(D) - \text{Info}_{\text{income}}(D) \\ &= 0.9183 - 0.8637 \\ &= 0.0546 \end{aligned}$$

student	$p_i$	$n_i$	$I(p_i, n_i)$
Yes	5	1	0.6500
No	3	3	1

$$I(5,1) = -\frac{5}{6} \log_2 \left( \frac{5}{6} \right) - \frac{1}{6} \log_2 \left( \frac{1}{6} \right) = 0.6500$$

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

$$\text{Info}_{\text{student}} = \frac{6}{12} I(5,1) + \frac{6}{12} I(3,3) = \frac{6}{12} (0.6500) + \frac{6}{12} (1) = 0.825 \neq$$

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.9183 - 0.825 = 0.0933 \#$$

$x_1$	$x_2$	$x_3$	$x_4$	$y$
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

credit_rating	$p_i$	$n_i$	$I(p_i, n_i)$
fair	6	1	0.5917
excellent	2	3	0.9710

$$I(6,1) = -\frac{6}{7} \log_2 \left( \frac{6}{7} \right) - \frac{1}{7} \log_2 \left( \frac{1}{7} \right) = 0.5917$$

$$I(2,3) = -\frac{2}{5} \log_2 \left( \frac{2}{5} \right) - \frac{3}{5} \log_2 \left( \frac{3}{5} \right) = 0.9710$$

$$\begin{aligned} \text{Gain}(\text{credit\_rating}) &= \text{Info}(D) - \text{Info}_{\text{credit\_rating}}(D) \\ &= 0.9183 - 0.7497 \\ &= 0.1686 \# \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{credit\_rating}}(D) &= \frac{7}{12} I(6,1) + \frac{5}{12} I(2,3) \\ &= \frac{7}{12} (0.5917) + \frac{5}{12} (0.9710) \\ &= 0.7497 \# \end{aligned}$$

$$\text{Gain}(\text{age}) = 0.1804 \text{ — 2nd best}$$

$$\text{Gain}(\text{income}) = 0.0546$$

$$\text{Gain}(\text{student}) = 0.0933$$

$$\text{Gain}(\text{credit\_rating}) = 0.1686$$

## 6.500 age node

age ≤ 30

income	student	credit_rating	buys computer
high	no	fair	no
high	no	excellent	no
low	yes	fair	yes
medium	yes	excellent	yes

$$Info(D, \leq 30) = I(2, 2) = -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right) = 1$$

income	p <sub>i</sub>	n <sub>i</sub>	I(p <sub>i</sub> ; n <sub>i</sub> )
high	0	2	0
medium	1	0	0
low	1	0	0

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

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

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

$$Info_{income}(D, \leq 30) = \frac{2}{4} I(0, 2) + \frac{1}{4} I(1, 0) + \frac{1}{4} I(1, 0) = \frac{2}{4}(0) + \frac{1}{4}(0) + \frac{1}{4}(0) = 0$$

$$Gain_{income} = Info(D, \leq 30) - Info_{income}(D, \leq 30) = 1 - 0 = 1$$

student	p <sub>i</sub>	n <sub>i</sub>	I(p <sub>i</sub> ; n <sub>i</sub> )
Yes	2	0	0
No	0	2	0

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

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

$$Info_{student}(D, \leq 30) = \frac{2}{4} I(2, 0) + \frac{2}{4} I(0, 2) = \frac{2}{4}(0) + \frac{2}{4}(0) = 0$$

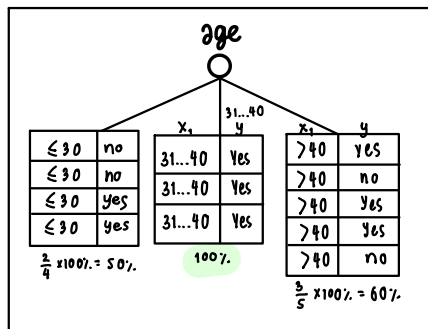
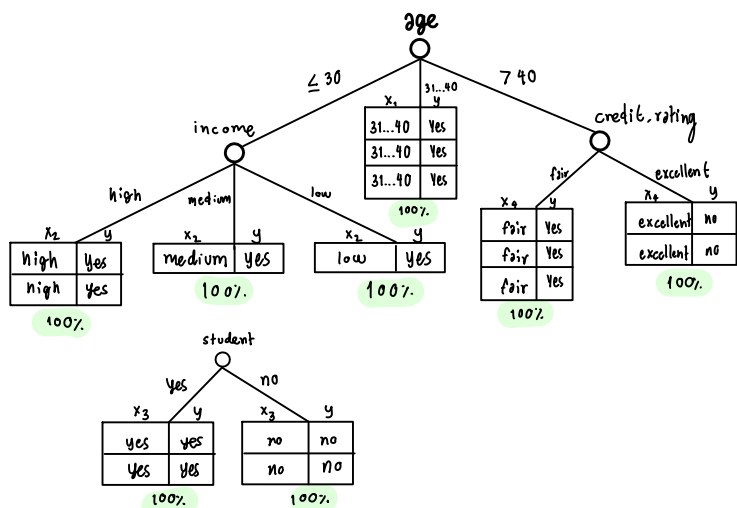
$$Gain_{student} = Info(D, \leq 30) - Info_{student}(D, \leq 30) = 1 - 0 = 1$$

credit_rating	p <sub>i</sub>	n <sub>i</sub>	I(p <sub>i</sub> ; n <sub>i</sub> )
Fair	1	1	1
excellent	1	1	1

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

$$Info_{credit\_rating}(D, \leq 30) = \frac{2}{4} I(1, 1) + \frac{2}{4} I(1, 1) = \frac{2}{4}(1) + \frac{2}{4}(1) = 1$$

$$Gain_{credit\_rating} = Info(D, \leq 30) - Info_{credit\_rating}(D, \leq 30) = 1 - 1 = 0$$



age > 40

income	student	credit_rating	buys computer
medium	no	fair	yes
low	yes	fair	yes
low	yes	excellent	no
medium	yes	fair	yes
medium	no	excellent	no

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

income	p <sub>i</sub>	n <sub>i</sub>	I(p <sub>i</sub> ; n <sub>i</sub> )
high	0	0	0
medium	2	1	0.9183
low	1	1	1

$$I(2, 1) = -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) = 0.9183$$

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

$$Info_{income}(D, > 40) = \frac{3}{5} I(2, 1) + \frac{2}{5} I(1, 1) = \frac{3}{5}(0.9183) + \frac{2}{5}(1) = 0.9326$$

$$Gain_{income} = Info(D, > 40) - Info_{income}(D, > 40) = 0.971 - 0.9326 = 0.0384$$

student	p <sub>i</sub>	n <sub>i</sub>	I(p <sub>i</sub> ; n <sub>i</sub> )
Yes	2	1	0.9183
No	1	1	1

$$I(2, 1) = -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) = 0.9183$$

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

$$Info_{student}(D, > 40) = \frac{3}{5} I(2, 1) + \frac{2}{5} I(1, 1) = \frac{3}{5}(0.9183) + \frac{2}{5}(1) = 0.9326$$

$$Gain_{student} = Info(D, > 40) - Info_{student}(D, > 40) = 0.971 - 0.9326 = 0.0384$$

credit_rating	p <sub>i</sub>	n <sub>i</sub>	I(p <sub>i</sub> ; n <sub>i</sub> )
Fair	3	0	0
excellent	0	2	0

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

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

$$Info_{credit\_rating}(D, > 40) = \frac{3}{5} I(3, 0) + \frac{2}{5} I(0, 2) = \frac{3}{5}(0) + \frac{2}{5}(0) = 0$$

$$Gain_{credit\_rating} = Info(D, > 40) - Info_{credit\_rating}(D, > 40) = 0.971 - 0 = 0.971$$

