

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

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

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

$$\begin{aligned} \text{Info}(D) &= I(9, 5) = -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) \\ &= 0.940 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{age}}(D) &= \frac{5}{14} I(2, 3) + \frac{4}{14} I(4, 0) + \frac{5}{14} I(3, 2) \\ &= 0.644 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \frac{4}{14} I(2, 2) + \frac{6}{14} I(4, 2) + \frac{4}{14} I(3, 1) \\ &= \frac{4}{14} \left(-\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) \right) + \frac{6}{14} \left(-\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) \right) \\ &\quad + \frac{6}{14} \left(-\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) \right) \\ &= 0.911 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{student}}(D) &= \frac{7}{14} I(6, 1) + \frac{7}{14} I(4, 3) \\ &= \frac{7}{14} \left(-\frac{6}{7} \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \log_2\left(\frac{1}{7}\right) \right) + \frac{7}{14} \left(-\frac{4}{7} \log_2\left(\frac{4}{7}\right) - \frac{3}{7} \log_2\left(\frac{3}{7}\right) \right) \\ &= 0.788 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{credit_rating}}(D) &= \frac{3}{14} \left(-\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) \right) + \frac{6}{14} \left(-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right) \\ &= \frac{3}{14} I(6, 2) + \frac{6}{14} I(3, 2) \\ &= 0.892 \end{aligned}$$

$$\text{Gain}_{\text{age}} = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.940 - 0.644 = 0.296$$

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

$$\text{Gain}_{\text{student}} = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.940 - 0.788 = 0.152$$

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

Root node = di Gain n'w n'g n' o'p age d'di Gain $\rightarrow 0.246$

if age ≤ 30

Info age ≤ 30 (D) = I(2, 3)

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

$$= 0.971$$

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

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

$$= 0.4$$

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

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

$$= 0$$

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

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

$$= 0.951$$

Gain (income) = Info age ≤ 30 (D) - Info income (D) = $0.971 - 0.400 = 0.571$

Gain student = Info age ≤ 30 (D) - Info student (D) = $0.971 - 0 = 0.971$

Gain credit-rating = Info age ≤ 30 (D) - Info credit-rating (D) = $0.971 - 0.951 = 0.020$

decision node n'w n' student n'w n' can n'g n'.

Info age = 31...40 (D) = I(4, 0)

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

$$= 0$$

31...40 n'w n' decision node n'w n' n'w n' I(4, 0) n'w n' n'w n' age 31...40 n'w n' n'w n'.

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

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

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

$$= 0.951$$

Info student (D) = $\frac{2}{5} I(2, 1) + \frac{3}{5} I(1, 1) = \frac{2}{5} \left(-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \right) + \frac{3}{5} \left(-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right)$

$$= 0.951$$