จงคำนวณหา

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

จากสูตร

❏ Expected information (entropy) needed to classify a tuple in D:

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

❏ Information needed (after using A to split D into v partitions) to classify D:

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

❏ Information gained by branching on attribute A

$$Gain(A) = Info(D) - Info_A(D)$$

หา Info (D)

$$Info(D) = I(9,5) = -\frac{9}{14} \log_{(2)}\left(\frac{9}{14}\right) - \frac{5}{14} \log_{(2)}\left(\frac{5}{14}\right)$$

$$= 0.94$$

หา $Info_{age}(D)$

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

✱หาแยก $I(2,3) = -\frac{2}{5} \log_{(2)}\left(\frac{2}{5}\right) - \frac{3}{5} \log_{(2)}\left(\frac{3}{5}\right) = 0.971$

$$I(4,0) = -\frac{4}{4} \log_{(2)}\left(\frac{4}{4}\right) - \frac{0}{4} \log_{(2)}\left(\frac{0}{4}\right) = 0$$

$$I(3,2) = -\frac{3}{5} \log_{(2)}\left(\frac{3}{5}\right) - \frac{2}{5} \log_{(2)}\left(\frac{2}{5}\right) = 0.971$$

แทนค่า จะได้ $Info_{age}(D) = \frac{5}{14}(0.971) + \frac{4}{14}(0) + \frac{5}{14}(0.971) = 0.694$

หา Gain (age)

$$Gain(age) = 0.94 - 0.694 = 0.246$$

หา Info$_{income}$ (D)

$$\text{Info.}_{income}(D) = \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1)$$

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

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

$$I(3,1) = -\frac{3}{4}\log_{(2)}\left(\frac{3}{4}\right) - \frac{1}{4}\log_{(2)}\left(\frac{1}{4}\right) = 0.811$$

แทนค่า จะได้ $\text{Info}_{income}(D) = \frac{4}{14}(1) + \frac{6}{14}(0.918) + \frac{4}{14}(0.811)$

$$= 0.911$$

หา Gain (income)

$$\text{Gain (income)} = 0.94 - 0.911 = 0.029$$

หา Info$_{student}$ (D)

$$\text{Info}_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4)$$

* หาจาก $I(6,1) = -\frac{6}{7}\log_{(2)}\left(\frac{6}{7}\right) - \frac{1}{7}\log_{(2)}\left(\frac{1}{7}\right) = 0.592$

$$I(3,4) = -\frac{3}{7}\log_{(2)}\left(\frac{3}{7}\right) - \frac{4}{7}\log_{(2)}\left(\frac{4}{7}\right) = 0.985$$

แทนค่า จะได้ $\text{Info}_{student}(D) = \frac{7}{14}(0.592) + \frac{7}{14}(0.985)$

$$= 0.789$$

หา Gain (student)

$$\text{Gain (student)} = 0.94 - 0.789$$

$$= 0.151$$

หา $Info_{credit\_rating}(D)$

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

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

$$I(3,3) = -\frac{3}{6}\log_{(2)}\left(\frac{3}{6}\right) - \frac{3}{6}\log_{(2)}\left(\frac{3}{6}\right) = 1$$

แทนค่า จะได้ $Info_{credit\_rating}(D) = \frac{8}{14}(0.811) + \frac{6}{14}(1) = 0.892$

หา Gain (credit_rating)

$$Gain(credit\_rating) = 0.94 - 0.892 = 0.048$$

จากการหา Gain ทั้งหมด จะได้

Gain (age) $= 0.246$

Gain (income) $= 0.029$

Gain (student) $= 0.151$

Gain (credit_rating) $= 0.048$

ให้เลือก Gain ที่มีค่ามากที่สุด มาพิจารณาเป็นส่วนแรก จะได้ดังนี้
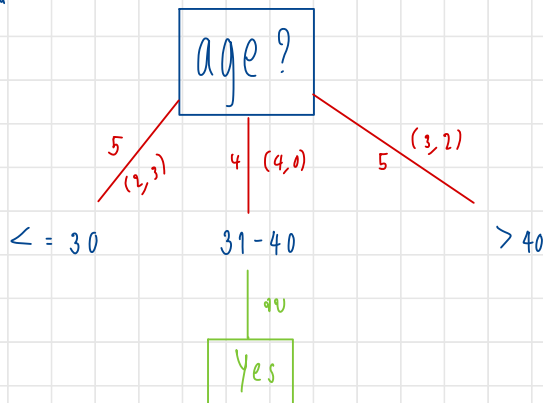
age ($\angle = 30$)

หา Info(D) ของ age ($\angle = 30$) จะได้

$$\text{Info}(D) = I(2,3) = 0.971$$

หา $\text{Info}_{income}(D)$ ของ age จะได้

$$\text{Info}_{income}(D) \text{ ของ age } (\angle = 30) = \frac{2}{5} I(0,2) + \frac{2}{5} I(1,1) + \frac{1}{5} I(1,0)$$

✳ หาแยก $I(0,2) = -\frac{0}{2}\log_{(2)}\left(\frac{1}{2}\right) - \frac{2}{2}\log_{(2)}\left(\frac{2}{2}\right) = 0$

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

$$I(1,0) = -\frac{1}{1}\log_{(2)}\left(\frac{1}{1}\right) - \frac{0}{1}\log_{(2)}\left(\frac{0}{2}\right) = 0$$

แทนค่า $\text{Info}_{income}(D)$ ของ age ($\angle = 30$) $= \frac{2}{5}(0) + \frac{2}{5}(1) + \frac{1}{5}(0) = 0.4$

หา Gain (income) ของ age ($\angle = 30$) จะได้

$$\text{Gain (income) ของ age } (\angle = 30) = 0.971 - 0.4 = 0.571$$

หา $\text{Info}_{student}(D)$ ของ age ($\angle = 30$)

$$\text{Info}_{student}(D) \text{ ของ age } (\angle = 30) = \frac{2}{5} I(2,0) + \frac{3}{5} I(0,3)$$

สิ้นสุด   Yes → Yes (buy_computer), No → no (buy_computer)

เลือกแบ่งด้วย student เพราะสามารถแบ่งข้อมูลต่างๆได้สมบูรณ์

age (>40)

หา Info (D) ของ age (>40) จะได้
$$\text{Info (D) ของ age } (>40) = I(3,2) = 0.971$$

หา Info_{income} (D) ของ age (>40)
$$\text{Info}_{income} \text{ (D) ของ age } (>40) = \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1)$$

※ หาแยก $I(2,1) = -\frac{2}{3} \log_{(2)} \left(\frac{2}{3}\right) - \frac{1}{3} \log\left(\frac{1}{3}\right) = 0.918$

$\quad\quad\quad I(1,1) = 1$

แทนค่า จะได้ Info_{income} (D) ของ age $(>40) = \frac{3}{5}(0.918) + \frac{2}{5}(1) = 0.951$

หา Gain (income) ของ age (>40)
$$\text{Gain (income) ของ age } (>40) = 0.971 - 0.951 = 0.02$$

หา $\text{Info}_{student}$ (D) ของ age (>40)
$$\text{Info}_{student} \text{ (D) ของ age } (>40) = \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1)$$

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

$\quad\quad\quad I(1,1) = 1$

แทนค่า จะได้ $\text{Info}_{student}$ (D) ของ age $(>40) = \frac{3}{5}(0.918) + \frac{2}{5}(1) = 0.951$

หา Gain (student) ของ age (>40)
$$\text{Gain (student) age } (>40) = 0.971 - 0.951 = 0.02$$

หา $\text{Info}_{credit\_rating}$ (D) ของ age (>40)
$$\text{Info}_{credit\_rating} \text{ (D) ของ age } (>40) = \frac{3}{5} I(3,0) + \frac{2}{5} I(0,2)$$

สิ้นสุด fair → Yes (buy_computer), excellent → no (buy_computer)

เลือกแบ่งด้วย credit_rating เนื่องจากสามารถแบ่งข้อมูลต่างๆได้อย่างสมบูรณ์

จากคำตอบที่ได้จากการคำนวณทั้งหมด สรุปได้ดังนี้