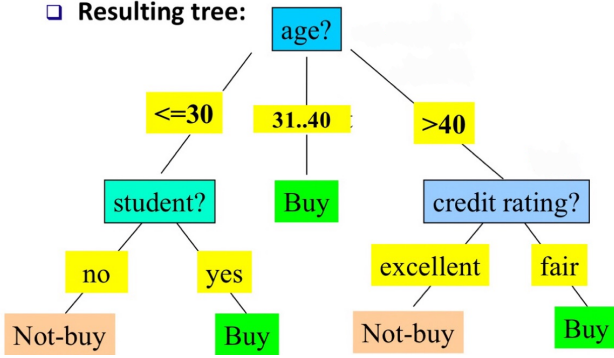


□ Resulting tree:



• Class P: buys_computer = "yes" = 9

• Class N: buys_computer = "No" = 5

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

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

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

2.1) $Info_{age}(D) = \frac{5}{14} I(2,9) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2)$

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

$$= 0.694 \#$$

2.2) $Info_{income}(D) = \frac{4}{14} I(2,1) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1)$

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

$$= 0.911 \#$$

2.3) $Info_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4)$

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

$$= 0.789 \#$$

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

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

$$= 0.892 \#$$

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

3.คำนวณ Information Grain โดยค่า Grain ที่สูงที่สุดเป็น root node

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

$$3.1) \text{Gain}(\text{age}) = 0.940 - 0.694 = 0.246$$

$$3.2) \text{Gain}(\text{income}) = 0.940 - 0.911 = 0.029$$

$$3.3) \text{Gain}(\text{student}) = 0.940 - 0.789 = 0.151$$

$$3.4) \text{Gain}(\text{credit_rating}) = 0.940 - 0.892 = 0.048$$

∴ เลือก Gain(age) เป็น root node เนื่องจากมีค่า Gain สูงที่สุด #

4. แยกกลุ่มของ Feature ตามค่าใน root node

4.1) ≤ 30

age	income	student	credit_rating	buys_computer
≤ 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

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

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

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

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

4.2) 31...40

age	income	student	credit_rating	buys_computer
31...40	high	no	fair	yes
31...40	low	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes

คำนวณ Information Grain กลุ่ม age ≤ 30

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

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

$$\text{Gain}(\text{credit_rating}) = 0.971 - 0.951 = 0.02$$

∴ เลือก Gain(student) เป็น root note ในกลุ่ม Age ≤ 30 #

จาก yes = 4

No = 0

ถ้า age 31...40 สามารถตอบว่า yes ใน buys_computer ได้เลย #

4.3) > 40

age	income	student	credit_rating	buys_computer
> 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

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

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

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

$$\text{Info}_{\text{credit_rating}}(D) = \frac{3}{6} I(1,0) + \frac{2}{6} I(0,2) = 0$$

คำนวณ Information Grain กลุ่ม age > 40

$$\text{Gain}(\text{income}) = 0.971 - 0.951 = 0.02$$

$$\text{Gain}(\text{student}) = 0.971 - 0.951 = 0.02$$

$$\text{Gain}(\text{credit_rating}) = 0.971 - 0 = 0.971$$

∴ เลือก Gain(credit_rating) เป็น root note ในกลุ่ม Age > 40 #

5. สร้าง Decision tree

