

Link code:

[https://colab.research.google.com/drive/1pBkW3E3WyXC\\_ihRH5vVC3Ru4Hrdce-OT#scrollTo=pYwvQVfRQQpR](https://colab.research.google.com/drive/1pBkW3E3WyXC_ihRH5vVC3Ru4Hrdce-OT#scrollTo=pYwvQVfRQQpR)

Bài làm

Thứ      ngày      tháng      năm

ID3

IE tổng quát  $(S, 9^+, 5^-)$

$$= -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) \approx 0,94$$

\* Outlook

$$IE(S, \text{Sunny}, 2^+, 3^-) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \approx 0,97$$
$$IE(S, \text{Rain}, 3^+, 2^-) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \approx 0,97$$
$$\Rightarrow IG = 0,94 - \frac{5}{14} \cdot 0,97 - \frac{5}{14} \cdot 0,97 = 0,247$$

\* Temperature

$$IE(S, \text{Hot}, 2^+, 2^-) = -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) = 1$$
$$IE(S, \text{Mild}, 4^+, 2^-) = -\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) = 0,918$$
$$IE(S, \text{Cool}, 3^+, 1^-) = -\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) = 0,811$$
$$\Rightarrow IG = 0,94 - \frac{14}{14} \cdot 1 - \frac{6}{14} \cdot 0,918 - \frac{4}{14} \cdot 0,811$$
$$\approx 0,029$$

★ STARBOOK

\* Humidity

$$IE(S, \text{High}, 3^+, 4^-) = -\frac{3}{7} \log_2 \frac{3}{7} - \frac{4}{7} \log_2 \frac{4}{7} \\ = 0,985$$

$$IE(S, \text{Normal}, 6^+, 1^-) = -\frac{6}{7} \log_2 \frac{6}{7} - \frac{1}{7} \log_2 \frac{1}{7} \\ \approx 0,592$$

$$\Rightarrow IG = 0,94 - \frac{7}{14} \cdot 0,985 - \frac{7}{14} \cdot 0,592 = 0,152$$

\* Wind

$$IE(S, \text{Weak}, 6^+, 2^-) = -\frac{6}{8} \log_2 \frac{6}{8} - \frac{2}{8} \log_2 \frac{2}{8} \\ = 0,811$$

$$IE(S, \text{Strong}, 3^+, 3^-) = -\frac{3}{6} \log_2 \frac{3}{6} - \frac{3}{6} \log_2 \frac{3}{6} \\ = 1$$

$$\Rightarrow IG = 0,94 - \frac{8}{14} \cdot 0,811 - \frac{6}{14} \cdot 1 = 0,048$$

\* Sunny :  $IE = 0,97$

\* Temperature :  $IE(S, \text{Hot}, 0^+, 2^-) = 0$

$$IE(S, \text{Wild}, 1^+, 1^-) = 1$$

$$IE(S, \text{Cold}, 1^+, 0^-) = 0$$

$$\Rightarrow IG = 0,97 - \frac{2}{5} = 0,57$$



Humidity

$$IE(S, \text{Hight}, 0^+, 3^-) = 0$$

$$IE(S, \text{Normal}, 2^+, 0^-) = 0$$

$$\Rightarrow IG = 0,97$$

$$\text{Wind: } IE(S, \text{Weak}, 1^+, 2^-) = 0,918$$

$$IE(S, \text{Strong}, 1^+, 1^-) = 1$$

$$\Rightarrow IG = 0,0202$$

$$* \text{Rain: } IE = 0,97$$

$$\text{Temperature: } IE(S, \text{Wild}, 2^+, 1^-) = 0,918$$

$$IE(S, \text{Cold}, 1^+, 1^-) = 1$$

$$\Rightarrow IG = 0,97 - \frac{3}{5} \cdot 0,918 - \frac{2}{5} \cdot 1 = 0,0202$$

Humidity

$$IE(S, \text{Hight}, 1^+, 1^-) = 1$$

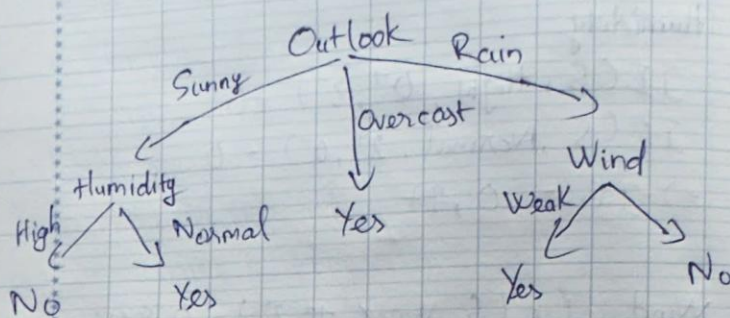
$$IE(S, \text{Normal}, 2^+, 1^-) = 0,918$$

$$\left. \begin{array}{l} IE(S, \text{Hight}, 1^+, 1^-) = 1 \\ IE(S, \text{Normal}, 2^+, 1^-) = 0,918 \end{array} \right\} IG = 0,0202$$

$$\text{Wind: } IE(S, \text{Weak}, 3^+, 0^-) = 0$$

$$IE(S, \text{Strong}, 0^+, 2^-) = 0$$

$$\left. \begin{array}{l} IE(S, \text{Weak}, 3^+, 0^-) = 0 \\ IE(S, \text{Strong}, 0^+, 2^-) = 0 \end{array} \right\} IG = 0,97$$



C4.5 Outlook

$$IG = 0,247$$

$$\text{Split}(S, \text{Outlook}) = -\frac{5}{14} \log_2 \frac{5}{14} - \frac{4}{14} \log_2 \frac{4}{14} - \frac{5}{14} \log_2 \frac{5}{14}$$

$$= 1,577$$

$$\text{Gain Ratio} = \frac{IG}{\text{Split}} = \frac{0,247}{1,577} = 0,157$$

Temperature :  $IG = 0,029$

$$\text{Split} = -\frac{4}{14} \log_2 \frac{4}{14} - \frac{6}{14} \log_2 \frac{6}{14} - \frac{4}{14} \log_2 \frac{4}{14}$$

$$= 1,557$$

$$\text{Gain Ratio} = \frac{IG}{\text{Split}} = \frac{0,029}{1,557} = 0,019$$

Humidity :  $IG = 0,152$  , Split = 1

$$\text{Gain Ratio} = 0,152$$



Wind:  $IG = 0,048$   
 $Split = 0,985$  }  $Gain Ratio = \frac{0,048}{0,985} = 0,049$

+ Outlook = Sunny

Temperature:  $IG = 0,57$ ,  $Split = 1,522$

Gain Ratio (GR) = 0,375

Humidity:  $IG = 0,97$ ,  $Split = 0,97$   
 $GR = 1$

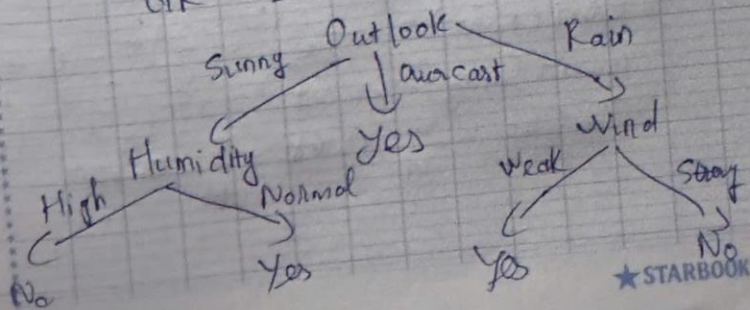
Wind:  $IG = 0,0202$ ,  $Split = 0,97$   
 $GR = 0,021$

+ Outlook = Rain

Temperature:  $IG = 0,0202$ ,  $Split = 0,97$  }  $GR = 0,021$

Humidity = Temperature

Wind:  $IG = 0,97$ ,  $Split = 0,97$   
 $GR = 1$



CART

$$\text{Gini} = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 = 0,459$$

Outlook

Sunny | Overcast, Rain (1)

Overcast | Sunny, Rain (2)

Rain | Sunny, Overcast (3)

$$\begin{aligned} 1) \text{ Gini (Sunny)} &= 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 \approx 0,48 \\ \text{Gini (Overcast, Rain)} &= 1 - \left(\frac{7}{9}\right)^2 - \left(\frac{2}{9}\right)^2 \approx 0,346 \\ \text{GiniSplit} &= \frac{5}{14} \cdot 0,48 + \frac{9}{14} \cdot 0,346 = 0,394 \end{aligned}$$

$$\begin{aligned} 2) \text{ Gini (Overcast)} &= 1 - \left(\frac{4}{4}\right)^2 - 0 = 0 \\ \text{Gini (Sunny, Rain)} &= 1 - \left(\frac{5}{10}\right)^2 - \left(\frac{5}{10}\right)^2 = 0,5 \\ \text{GiniSplit} &= \frac{4}{14} \cdot 0 + \frac{10}{14} \cdot 0,5 \approx 0,357 \end{aligned}$$

$$\begin{aligned} 3) \text{ Gini (Rain)} &= 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = 0,48 \\ \text{Gini (Sunny, Overcast)} &= 1 - \left(\frac{6}{9}\right)^2 - \left(\frac{3}{9}\right)^2 = 0,444 \\ \text{GiniSplit} &= \frac{5}{14} \cdot 0,48 + \frac{9}{14} \cdot 0,444 = 0,452 \end{aligned}$$



### Temperature

Hot | Mild, Cold (1)

Mild | Hot, Cold (2)

Cold | Hot, Mild (3)

$$1) \text{ Gini (Hot)} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0,5$$

$$\text{Gini (Mild, Cold)} = 1 - \left(\frac{3}{10}\right)^2 - \left(\frac{7}{10}\right)^2 = 0,42$$

$$\text{Gini Split} = \frac{4}{14} \cdot 0,5 + \frac{10}{14} \cdot 0,42 = 0,443$$

$$2) \text{ Gini (Mild)} = 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = 0,444$$

$$\text{Gini (Hot, Rain)} = 1 - \left(\frac{5}{8}\right)^2 - \left(\frac{3}{8}\right)^2 = 0,469$$

$$\text{Gini Split} = \frac{6}{14} \cdot 0,444 + \frac{8}{14} \cdot 0,469 = 0,457$$

$$3) \text{ Gini (Cold)} = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0,375$$

$$\text{Gini (Hot, Mild)} = 1 - \left(\frac{4}{10}\right)^2 - \left(\frac{6}{10}\right)^2 = 0,48$$

$$\text{Gini Split} = \frac{4}{14} \cdot 0,375 + \frac{10}{14} \cdot 0,48 = 0,45$$

### Humidity

$$\text{Gini (High)} = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 = 0,49$$

$$\text{Gini (Normal)} = 1 - \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2 = 0,245$$

$$\text{Gini Split} = \frac{7}{14} \cdot 0,49 + \frac{7}{14} \cdot 0,245 = 0,368$$

Wind

$$\text{Gini (Weak)} = 1 - \left(\frac{2}{8}\right)^2 - \left(\frac{6}{8}\right)^2 = 0,375$$

$$\text{Gini (Strong)} = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0,5$$

$$\text{Gini Split} = \frac{8}{14} \cdot 0,375 + \frac{6}{14} \cdot 0,5 = 0,422$$

