

We want to split first on the variable which maximizes the information gain

$$\text{Entropy}(\text{Pass}) = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \approx 0.918$$

Studied

F

1+, 2-

$$\begin{aligned} E &= -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \\ &= -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3}\right) \\ &\approx 0.918 \end{aligned}$$

T

3+, 0-

$$\begin{aligned} E &= -\frac{3}{3} \log_2 \frac{3}{3} - 0 \\ &= 0 \end{aligned}$$

GPA	Studied	Pass
L	F	F
L	T	T
M	F	F
M	T	T
H	F	T
H	T	T

$$\text{Gain}(S, \text{Studied}) = 0.918 - \frac{3}{6} \times 0.918 - \frac{3}{6} \times 0 = 0.459$$

GPA

L

1+, 1-

$$\begin{aligned} E &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \\ &= -\log_2 \frac{1}{2} \\ &= 1 \end{aligned}$$

M

1+, 1-

$$\begin{aligned} E &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \\ &= -\log_2 \frac{1}{2} \\ &= 1 \end{aligned}$$

H

2+, 0-

$$\begin{aligned} E &= -\frac{2}{2} \log_2 \frac{2}{2} - 0 \\ &= 0 \end{aligned}$$

$$\text{Gain}(S, \text{GPA}) = 0.918 - \frac{2}{6} \times 1 - \frac{2}{6} \times 1 - \frac{2}{6} \times 0 = 0.918 - \frac{4}{6} = 0.918 - 0.667 = 0.251$$

