

Lab 06 - 6.2

Sarah Griffioen
3/9/19

- Entropy of all data at parent node

12 instances of WillWait : 6 yes, 6 no

$$\text{Entropy}(\text{WillWait}) = - \left[\left(\frac{6}{12} \right) \log_2 \left(\frac{6}{12} \right) + \left(\frac{6}{12} \right) \log_2 \left(\frac{6}{12} \right) \right] \\ = 1$$

- Child's expected entropy for "hungry" split

Left child = "Yes"

7 instances of WillWait : 5 yes, 2 no

$$\text{Entropy}(\text{WillWait}) = - \left[\left(\frac{5}{7} \right) \log_2 \left(\frac{5}{7} \right) + \left(\frac{2}{7} \right) \log_2 \left(\frac{2}{7} \right) \right] \\ = 0.8631$$

Right child = "No"

5 instances of WillWait : 1 yes, 4 no

$$\text{Entropy}(\text{WillWait}) = - \left[\left(\frac{1}{5} \right) \log_2 \left(\frac{1}{5} \right) + \left(\frac{4}{5} \right) \log_2 \left(\frac{4}{5} \right) \right] \\ = 0.7219$$

$$\begin{aligned} & (\text{Remainder}) \\ \text{Entropy}(\text{Hungry}) &= \left(\frac{7}{12} \cdot 0.8631 \right) + \left(\frac{5}{12} \cdot 0.7219 \right) \\ &= 0.5035 + 0.3008 \\ &= 0.8043 \end{aligned}$$

- Information Gain

$$I(\text{parent}) - I(\text{child})$$

$$1 - 0.8043$$

$$\boxed{0.1957}$$

This question is better than "Type" since "Type" isn't even a part of the decision tree. It does not yield any information about the dataset. This question is worse than "Patrons" because "Patrons" gained 0.46 bits of information about the dataset while "Hungry" only gained 0.1957 bits of information.