

Decision Tree Entropy and Information Gain Calculations

Entropy Calculations

Entropy is computed using the formula:

$$H(X) = - \sum p_i \log_2 p_i$$

Entropy of Purchase Size = Small

From the table, the rows with **Purchase Size = Small**:

$$H(\text{Small}) = - \left(\frac{0}{3} \log_2 \frac{0}{3} + \frac{3}{3} \log_2 \frac{3}{3} \right) = 0$$

Entropy of Purchase Size = Medium

From the table, the rows with **Purchase Size = Medium**:

$$H(\text{Medium}) = - \left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3} \right)$$

Approximating logarithm values:

$$\log_2(1/3) \approx -1.585, \quad \log_2(2/3) \approx -0.585$$

$$H(\text{Medium}) \approx 0.918$$

Entropy of Purchase Size = Large

From the table, the rows with **Purchase Size = Large**:

$$H(\text{Large}) = - \left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right)$$

$$H(\text{Large}) \approx 0.918$$

Information Gain Calculation

The overall entropy before splitting:

$$H(\text{Total}) = - \left(\frac{3}{9} \log_2 \frac{3}{9} + \frac{6}{9} \log_2 \frac{6}{9} \right) = 0.918$$

The weighted entropy after splitting:

$$\begin{aligned} H_{\text{split}} &= \frac{3}{9} H(\text{Small}) + \frac{3}{9} H(\text{Medium}) + \frac{3}{9} H(\text{Large}) \\ &= \frac{3}{9}(0) + \frac{3}{9}(0.918) + \frac{3}{9}(0.918) = 0.612 \end{aligned}$$

$$\text{Information Gain} = H(\text{Total}) - H_{\text{split}}$$

$$= 0.918 - 0.612 = 0.306$$