Steps to Build a DT
A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous).

To build a decision tree, we need to calculate two types of entropy using frequency tables:
- a) Entropy using the frequency table of the target attribute.
- b) Joint Entropy using the frequency table (Target, other attribute)

**Step 1: Calculate Entropy of the Target (Measure the uncertainty associated with the target)**

\[
H(Y) = - \sum_{i=1}^{c} p(Y = i) \log_2 p(i)
\]

\[
H(Y) = -\frac{5}{8} \log_2 \left(\frac{5}{8}\right) - \frac{3}{8} \log_2 \left(\frac{3}{8}\right) = 0.95
\]
• **Step 2:** Calculate Conditional Entropy for the target and each feature (*measure the uncertainty associated with target given each feature*)
  
  - \( H(Y|X_1) \)
  - \( H(Y|X_2) \)
  
  Note that: \( H(Y|X_1) \leq H(Y) \)

• **Step 3:** Calculate Information Gain for each feature:

  Think of IG as “How much uncertainty remains in the target after removing the effect of the feature”

\[
IGain(Y, X_1) = H(Y) - H(Y|X_1)
\]
\[
IGain(Y, X_2) = H(Y) - H(Y|X_2)
\]

• **Step 4:** Choose attribute with the largest information gain as the decision node.
  
  - A branch with entropy of 0 is a leaf node.
  - A branch with entropy more than 0 needs further splitting.

• **Step 5:** Recourse on non-leaf branches until all data is classified