1. For the following data set, apply ID3 separately, and show all steps of derivation (computation, reasoning, developing / final decision trees, and rules).

	color	shape	size	class
1	red	square	big	+
2	blue	square	big	+
3	red	round	small	-
4	green	square	small	-
5	red	round	big	+
6	green	round	big	-

$$Entropy(t) = -\sum_{j} \mathrm{p}(j|t) \log_2 \mathrm{p}(j|t)$$

Here class is the target attribute and has two values (+ and -). So it is a binary classification problem.

For a binary classification problem

- If all examples are positive or all are negative then entropy will be ***zero*** i.e. low.
- If half of the examples are of positive class and half are of negative class then entropy is **one** i.e. high.

1. Calculating Initial Entropy

   Out of 6 instances, 3 are + and 3 are –

$$P(+) = -\left(\frac{3}{6}\right) * \log_2\left(\frac{3}{6}\right) = 0.5$$

$$P(-) = -\left(\frac{3}{6}\right) * \log_2\left(\frac{3}{6}\right) = 0.5$$

$$Entropy(t) = E(t) = 0.5 + 0.5 = 1$$

   Note: 1 indicates that the classes are highly impure. It is true in our case as there are equal number of observations with target class + and –

2. For every feature we will calculate entropy and information gain

   For attribute color

$$E(Color = red) = -\frac{2}{3} * \log_2\frac{2}{3} - \frac{1}{3} * \log_2\frac{1}{3} \approx 0.92$$

$$E(Color = blue) = -\frac{1}{1} * \log_2 \frac{1}{1} - 0 = 0$$

$$E(Color = green) = -0 - \frac{2}{2} * \log_2 \frac{2}{2} = 0$$

Average Entropy = $\frac{3}{6}(0.92) + \frac{1}{6}(0) + \frac{2}{6}(0) = 0.46$

Gain(Outlook) = $1 - 0.46 = \mathbf{0.54}$

For attribute shape

$$E(shape = square) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} \approx 0.92$$

$$E(shape = round) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0.92$$

Average Entropy = $\frac{3}{6}(0.92) + \frac{3}{6}(0.92) = 0.92$

Gain(Outlook) = $1 - 0.92 = \mathbf{0.08}$

For attribute size

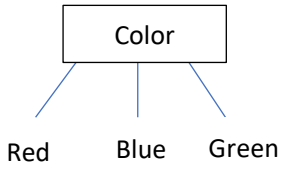
$$E(size = big) = -\frac{3}{4} * \log_2 \frac{3}{4} - \frac{1}{4} * \log_2 \frac{1}{4} \approx 0.81$$

$$E(size = small) = 0 - \frac{2}{2} * \log_2 \frac{2}{2} = 0$$

Average Entropy = $\frac{4}{6}(0.81) + \frac{2}{6}(0) = 0.54$

Gain(Outlook) = $1 - 0.54 = \mathbf{0.46}$

Feature ‘color’ provides more information on the ‘class’ as it has the highest information gain and hence will be chosen as the first splitting attribute



Likewise, we create the entire tree by selecting the splitting attribute as the attribute that gives the most information.