Information Gain Formula

Note that the child groups are weighted equally in this case since they're both the same size, for all splits. In general, the average entropy for the child groups will need to be a *weighted* average, based on the number of cases in each child group. That is, for m items in the first child group and n items in the second child group, the information gain is:

Information Gain =

$$\text{Entropy}(Parent) - \left[\frac{m}{m+n}\text{Entropy}(Child_1) + \frac{n}{m+n}\text{Entropy}(Child_2)\right]$$

Quiz Question

Assume we have a dataset with columns Car_Price and Car_Engine_Size and we want to recommend a specific car that is a good recommendation for the 25 age group. You calculate entropy and find that

Splitting by the Car_Price column gave us an information gain of 0.74
Splitting by the Car_Engine_Size column gave us an information gain of 0.61.

Which column will you split by? A. Car_Price

a. Car_Price
b. Car_Engine_Size