Information Gain Equation:

Importance = Original Entropy - New Entropy

Original Entropy $= H(\frac{a}{a+b+c+...}, \frac{b}{a+b+c+...}, \frac{c}{a+b+c+...}, ...)$

Where $a, b, c, ...$ correspond to the number of remaining examples in the training data with some outcome and $a + b + c + ...$ corresponds to the sum of all remaining examples in the training data at this point in the tree.

New Entropy $= \sum_{k=1}^{d} \frac{a_k+b_k+c_k+...}{a+b+c+...} H(\frac{a_k}{a_k+b_k+c_k+...}, \frac{b_k}{a_k+b_k+c_k+...}, \frac{c_k}{a_k+b_k+c_k+...}, ...)$

Where d corresponds to the number of return values of the question and $a_k, b_k, c_k, ...$ correspond to the number of each outcome of the examples in the k return value of the question. $a + b + c + ...$ is again the sum of all examples and $a_k + b_k + c_k + ...$ is the sum of all examples for the k return value of the question.

Entropy $= H(x_i) = -\frac{1}{log(n)} \sum_{i=1}^{n} P(x_i) log_2 P(x_i)$

From the wikipedia definition of entropy https://en.wikipedia.org/wiki/Entropy_(information_theory).

The $\frac{1}{log(n)}$ is added to scale branches that have eliminated an example outcome to to have the same max entropy as those that have a larger number of example outcomes remaining.