**Before looking into Information gain, we’ll see weighted entropy:**

So we see in previous lecture that for independent variable entropy was 0.94

Now we’ll find Entropy for Otlook feature, see it has 3 types of values: Sunny, Overcast, Rainy.

* So first we’ll split the dataset on the basis of Sunny, overcast and Rainy as we can see in below image.
* Now we’ll find entropy for dependent variable ie Play Tennis for each of the set.

Let’s look in D1 of sunny outlook, here we have P(yes) = 2/5 and P(no) = 3/5.

And therefore entropy for D1 ie E\_d1 calculated as 0.97

Similarly E\_d2 = 0 and E\_d3 = 0.97

* Now we find the Probability of sunny, overcast and Rainy in outlook.

P(sunny) = 5/14

P(overcast) = 4/14

P(Rainy) = 5/14

Now weighted Entropy or Entropy for outlook is calculated as:

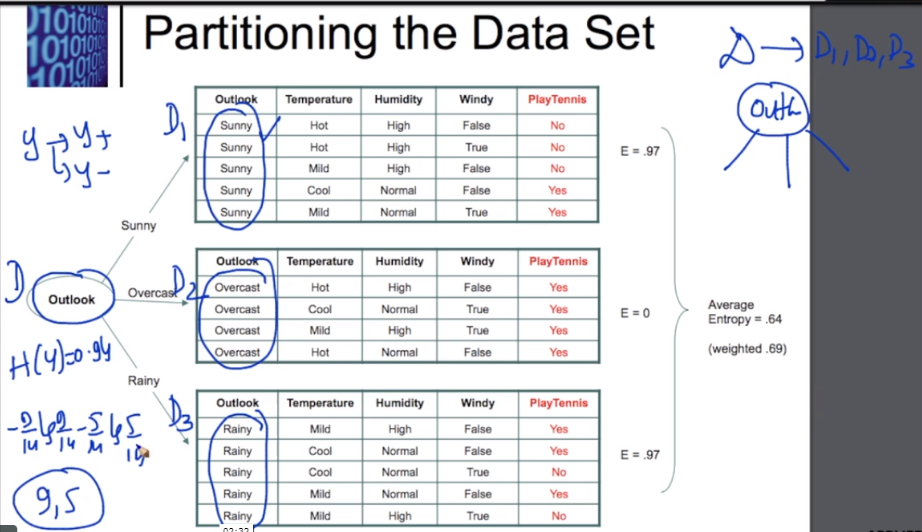
**Weighted Entropy** = P(sunny) \* E\_d1 + P(overcast) \* E\_d2 + P(Rainy) \* E\_d3

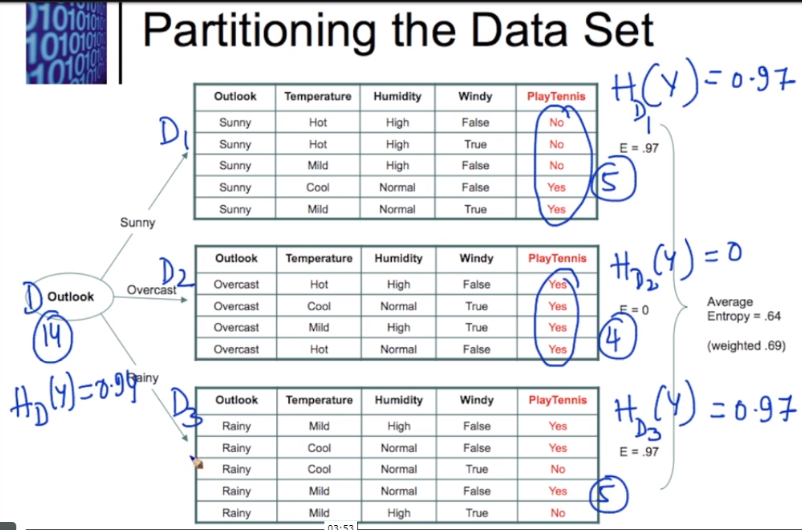
Now since we’ve find Weighted Entropy we can find Information Gain as:

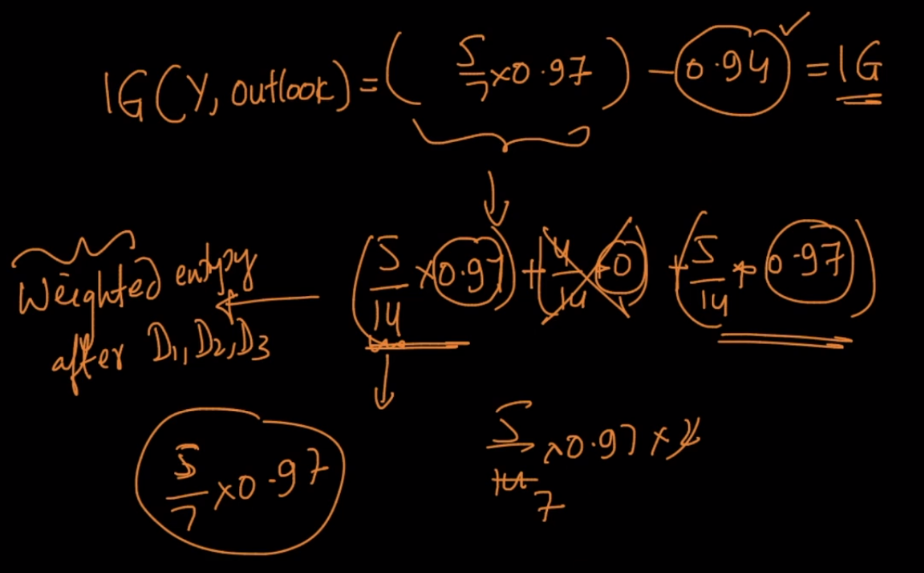
**Information Gain =** E(D) – Weighted Entropy

Here E(D) is the Entropy of the dependent feature over whole dataset, which we have calculated in pervious lecture.

**Information\_Gain = [Entropy(parent)] – [Weighted Average Entropy of child nodes]**

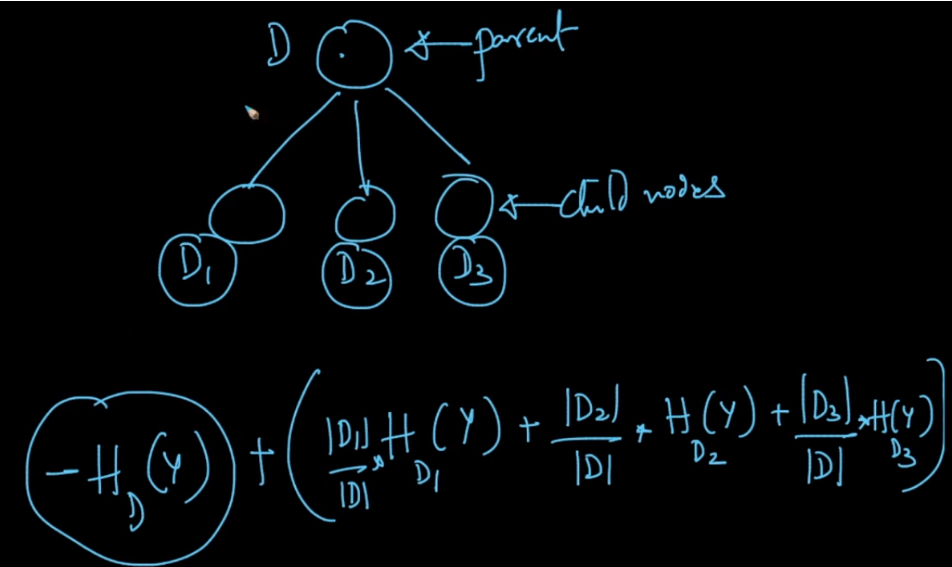






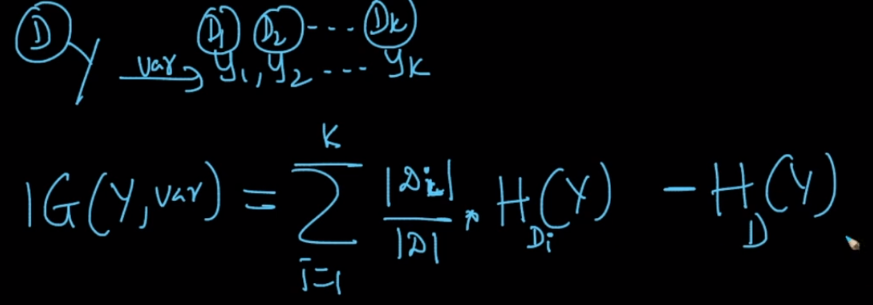
**How tree is calculated using Info gain:**

Whichever has the highest entropy will be choosen as parent node, now split can be done by that feature.



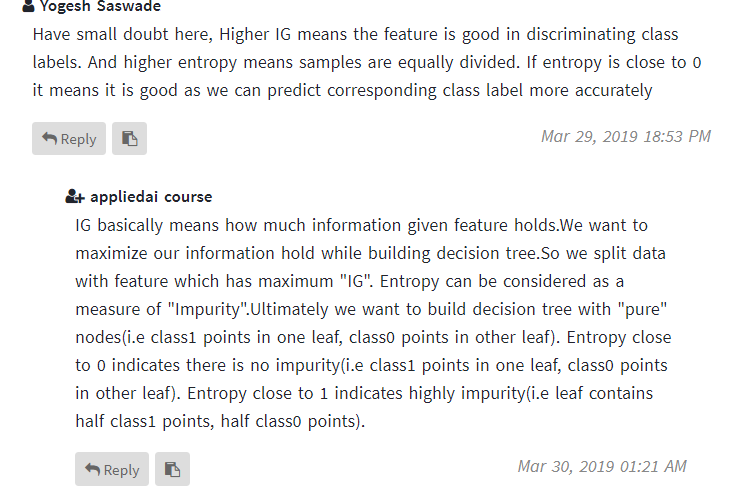
Generic Formula for Information gain can be written as

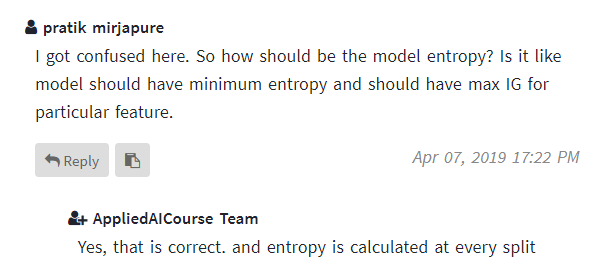
Suppose a feature Can have ‘K’ different values, then formula can be written as



Significane of Info gain: <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3067/building-a-decision-treeinformation-gain/4/module-4-machine-learning-ii-supervised-learning-models>

Must Read:





Comments:

