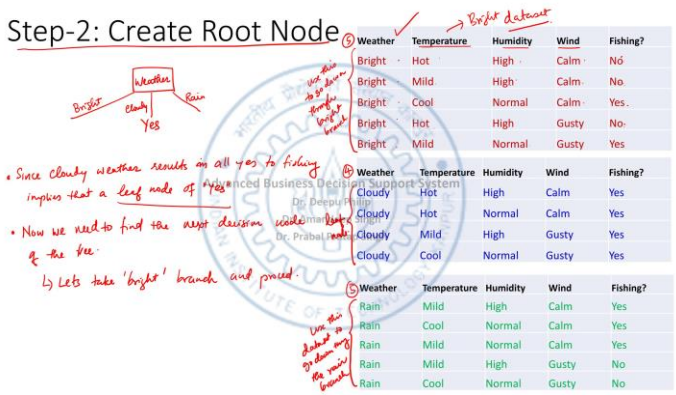
Advanced Business Decision Support Systems
Professor Deepu Philip
Department of Industrial Engineering and Management Engineering
Indian Institute of Technology, Kanpur
Professor Amandeep Singh
Imagineering Laboratory
Dr. Prabal Pratap Singh
Indian Institute of Technology, Kanpur
Lecture 15
Decision Tree Algorithm for Business Decision (Part 3 of 3)

Good afternoon everyone. Welcome to the fourth week lecture of the Business Decision Support System, the advanced course of the Web-Based Decision Support System under the NPTEL MOOC's program from IIT Kanpur. I am Dr. Deepu Philip and along with me Dr. Amandeep Singh Obreroi and Dr. Prabal Pratap Singh are teaching this course.

So, here we just finished the calculation of the Information Gain and we have decided that the weather will become the root node of the Decision Tree.



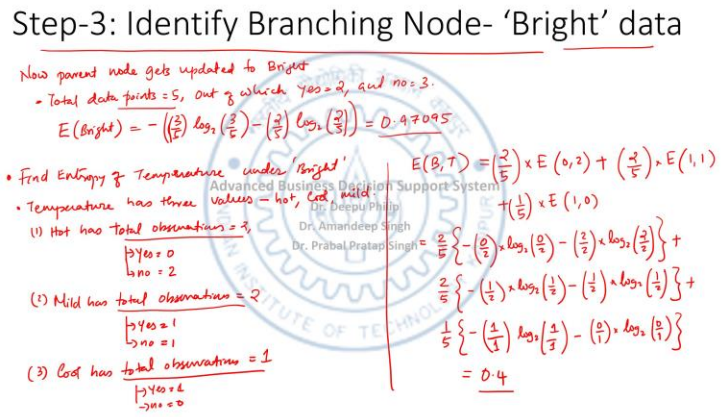
So, now let us create the root node. So, that is the next step. So, the Weather is the root node and you have 3 options in front of you that is bright. Then, the next one is Cloudy and the third one is Rain. So, if you look into this we have taken the entire data set. So, these 5 values here and 4 of them here. So, all the 14 observations is now splitted into 3 parts.

So, this part is the data set for the bright. So, if you want to go use this to go down through bright branch. You want to follow the bright branch of the tree, then you use this data set. And, second thing is that, Cloudy, if you look into it cloudy, all of the value is fishing. So, that means if you go to cloudy, that means, you will end up doing fishing.

So, then there is no need to for you here. So, we can say that since, cloudy weather results in all yes to fishing implies that leaf node. That means, if it is cloudy, there is no need to worry about any branching or anything. You need a leaf node, terminate there, that point itself. So, you will say yes.

So, if the weather is cloudy, you go for fishing, no matter what that is clarified. Now, the thing is so cloudy is done it is a now leaf node . If you want to go with the rain again, there is yes and no. So, use this data set to go down through the rain branch. So, the tree has now two branches, open bright and the rain and use this first data set for rain a bright and then, the other data set for the rain.

Now, we need to find the next decision node of the tree. So, here let us take bright branch and proceed. So, we are now going to the next branch, that is part of this. So, the step 3 is, identifying the branching node the bright. So, remember if you are going to do with the bright, this is the data set we are going to do. So, we will call the bright data set. So, we are no longer dealing with the entire 14 data set. Now, we are dealing with 5 of these.



So, now parent node gets updated to bright. So, these are 5 data points out of which yes equal 2, so that means, no equal to 3. So, now what we do is, we need to find the entropy of bright.

$$E(Bright) = -\left(\frac{3}{5}\right) log2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) log2\left(\frac{2}{5}\right)$$

So, if you do that, so just to easily make the calculation done let us just get back to the bright data set. So, the first one we are supposed to do is, this 3 by 5 and 2 by 5.

So, this is if you see is, the 3 by 5 is the 2 by 5 as per this numbers 3 by 5 and 2 by 5. So,

the first thing is take the log of 3 by 5 product of them. So, we get the entropy of that to be 0.97095 .

I am not giving you all the details of the calculation because we already gone through an extensive set through the excel. So, this might be, so I will just show it to you that is it. Now, from there find entropy of so now, in the data set. Now, you have temperature humidity and wind because you already made the decision on weather. So, you just need to focus on those 3 entropy of temperature under bright is the branch of the weather .

So, as we see again, temperature has 3 values hot mild and cool temperature has 3 values hot, cool and mild. Now, number 1 hot has, how many hot values are there. Hot is 1 and 2, the 2 values of hot has total observations equal to 2. So, yes is equal to and no is equal to so in the data set, the hot there is a no, hot there is a no yes with the hot. So, the 0 and no is 2.

Same way, number 2 mild has total observations equals how many are there mild 1 and 2 mild. So, 2 and how many of them are yes and no yes no. So, the mild no mild yes, so 1 yes and 1 no. The third one is cool has total observations equals how many observations are there in the cool? There is only one observation of cool that is yes. So, yes equal to 1 no is equal to 0.

$$E(B,T) = \left(\frac{2}{5}\right) * E(0,2) + \left(\frac{2}{5}\right) * E(1,1) + \left(\frac{1}{5}\right) * E(1,0)$$

$$\frac{2}{5}\left\{-\left(\frac{0}{2}\right) * log2\left(\frac{0}{2}\right) - \left(\frac{2}{2}\right) * log2\left(\frac{2}{2}\right)\right\} + \frac{2}{5}\left\{-\left(\frac{1}{2}\right) * log2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) * log2\left(\frac{1}{2}\right)\right\} + \frac{1}{5}\left\{-\left(\frac{1}{1}\right) * log2\left(\frac{1}{1}\right) - \left(\frac{0}{1}\right) * log2\left(\frac{0}{1}\right)\right\}$$

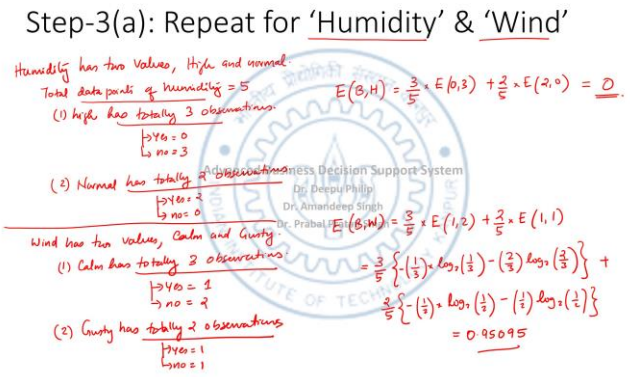
So, you have the 3 terms weighted out. So, if you look into this, the value that you want to calculate, you can calculate that value to be. So, let me explain that calculation keep this and we go to bright.

So, the first one is your 2 by 5, 2 by 5, 2 by 5 with 2 by 5, 2 by 5 and 1 by 5. So, this is the 2 by 5, 2 by 5, 1 by 5. So, those 3 values are there then the thing is the all we do is the 1 by 2. So, the first one is 0, this one is also typically. So, because this is 0 by 2, this whole term becomes 0 and log 1 will also go to 0.

So, this first term will become 0 the second time is 1 by 2. So, we take 1 by 2 then log 2 to the base of 2. This 1 is calculated from this number, then you multiply these 2 numbers D6 and C6. So, you get that product, you sum them up, then you multiply by that weight which is A6 0.4 is because it is 2 by 5.

And, then the last term here is this again 0 by 1. So, this is also 0. So, this total weight that comes out of this the information gain that comes out of this total thing will come to 0.4 because as the only one value that we get out of the this 1 0.4.

So, we write that here equals 0.4. That is the entropy of the weighted average entropy of the this entire.



Now, we do is, we go to repeat the same process for humidity and wind. So, let us do humidity has 2 values, high and normal.

So, again as I said earlier, the number of data sets is 5 of them and humidity, we are talking about high and normal. So, humidity has high and normal and total data points of humidity is equal to 5. Number 1 high has how many of them are high? High has 1, 2, 3. High has totally 3 observations. The total 3 observations yes and no that is what we need to find out.

So, high no high no high no. So, all the 3 of the high are no. So, high has 3 no and 0. Number 2, the normal humidity has totally how many observations? 2 observations, then you have yes equal to no is equal to. So, if you look at the yes normal yes normal yes. So, both of them are yes and 0 no.

$$E(B,H) = \frac{3}{5} * E(0,3) + \frac{2}{5} * E(2,0) = 0$$

So, remember the minute you see this 0 numbers, you know that it is going to be 0 and 0.

So, now let us take calm has totally how many data points calm 1, 2, 3, 3 data points, totally 3 observations. So, if it has 3 observations, how many of them are yes, how many of them are no. So, calm no calm yes. So, 1 yes and 2 no, this 1 yes and 2 no. Now, since there are totally 5, 2 was there already so that, 3 is already there that means, there should be 2 observations.

So, the yes is equal to and no is equal to. So, gusty we need, to go back gusty is a no and gusty 1 is an yes. So, we have 1 yes 1 no.

$$E(B,W) = \frac{3}{5} * E(1,2) + \frac{2}{5} * E(1,1)$$

$$\frac{3}{5}\left\{-\left(\frac{1}{3}\right) * log2\left(\frac{1}{3}\right) - \left(\frac{2}{3}\right) * log2\left(\frac{2}{3}\right)\right\} + \frac{2}{5}\left\{-\left(\frac{1}{2}\right) * log2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) * log2\left(\frac{1}{2}\right)\right\}$$

So, how do we calculate this? I have shown this earlier also, but just for the sake of calculation, we can do this quickly.

So, the easiest way to do it is, you first calculate the fraction, the 2 fractions 3 by 5 and 2 by 5. So, that is equals 3 by 5 equals 2 by 5 for the fractions are written out. In the next one, we need to do is, we calculate the thing inside the bracket is 1 by 3 and 2 by 3 that is the first term. So, that is equals 1 by 3 equals 2 by 3. And, then what you do is, you find the log of these 2 numbers log comma 2 equals log number comma 2.

Now, you do is, you do the product of these numbers times this and same thing gets repeated, here you have 2 values. So, the next thing you do is, you sum them this plus this got that value. So, then the next option is, you basically do the weighted value, this one multiplied by the weight. So, that gives you the first term inside this bracket, this is the first term. Second one is 1 by 2 and 1 by 2 and multiplied by 2 by 5.

So, it equals 1 by 2 equals 1 by 2, you take the log of this, it will come to minus 1, but that is ok, we still do that we know that comes here repeat the process is, when you have to chord this, you have to chord it exactly like this in the computer program. It is very difficult for you to do it manually. So, equals you sum them and then in this case is, you take this value and multiplied with the product. So, you get these 2 numbers. So, the total will be sum of this plus this which is equal to 0.95098. So, that is the value we get. So, we go to this one and we put it in the full screen and say it as 0.95095. So, this is the value the other case we got it as 0. So, both of them is done.

Step-3(b): Calculate IG for 'Bright'

- $IG_1(B,T) = E(Bright) - E(Temp) = 0.97095 - 0.4 = 0.57095$
- $IG_1(B,H) = E(Bright) - E(Humidity) = 0.97095 - 0 = 0.97095$ ✓
- $IG_1(B,N) = E(Bright) - E(Wind) = 0.97095 - 0.95095 = 0.02$

- Since humidity has the largest information gain, the next node becomes Humidity.
  - If Humidity = High, all data indicates no, implying no fishing will occur.
  - If Humidity = Normal, all data indicates yes, implying fishing will occur.

Now, what do we do is, calculate information gain for the bright. So, first thing is,

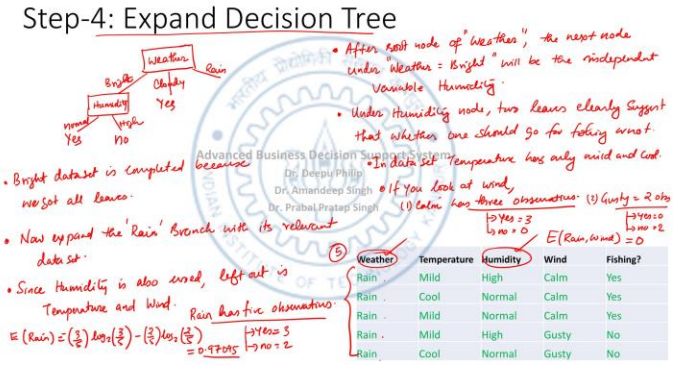
$$IG(B,T) = E(Bright) - E(Temp.) = 0.97095 - 0.4 = 0.57095$$

$$IG(B,H) = E(Bright) - E(Humidity) = 0.97095 - 0 = 0.97095$$

$$IG(B,W) = E(Bright) - E(Wind) = 0.97095 - 0.95095 = 0.02$$

So, in this you can see that this is the largest information gain. So, since humidity has the largest Information Gain the next node becomes humidity.

So, that means, what does that implies. So, we look at the data, if humidity is high, all the high values has no with it all the normal is yes. So, if we say, if humidity equals high, all data indicates no, implying no fishing will occur. If humidity equals normal, all data indicates yes. Yes means implying fishing will occur. Here, there is no fishing and here fishing will occur. So, that is the idea.



So, if that is the case, then what we do, we expand the Decision Tree. So, the Decision Tree so far, what we have written is, we have a Decision Tree of Weather and we had Bright and we had Cloudy and we had Rain and in Bright, now we said that, the next branch is Humidity. So, we make the next node humidity because that has the information gain and humidity has two branches normal and high. If it is normal, it is yes because we have seen that normal it is all yes.

So, you will go for fishing, it is high, it is no. So, you will not go for fishing. So, now that part is over. So, that means, you have pretty much completed using this data set and this is done. So, we can say that, the bright data set is completed because we got all leaves, we have definite decision coming out of it.

So, the cloudy again is also yes there is no need to branch. So, we have seen in our tree that if you see the weather is cloudy, go for fishing if the weather is bright, the check is the humidity is normal, go for fishing, humidity is high, do not go for fishing. So, we have completed that part of the tree we got all leaves. So, right here after root node of weather the next node under weather equal to bright will be the independent variable humidity. So, the decision is the dependent variable yes or no, but the independent variable the next one will be humidity.

So, now you have under humidity node. Under humidity node, two leaves clearly suggest that, whether one should go for fishing or not. So, that is the two aspects that we see there. Now, the only thing that is left out in front of us is, now expand the rain branch with its relevant data set. So, this is the rain data set. Now, the relevant data set 1, 2, 3, 4, 5, there are totally 5 observations.

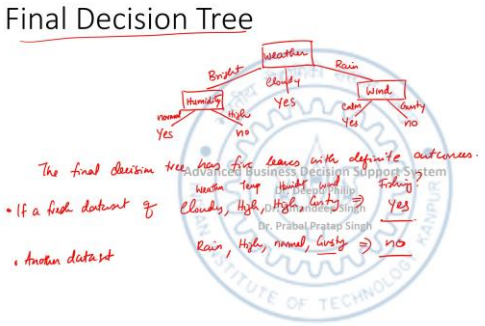
So, we have already consumed weather this is already done the root node and the humidity is also taken care of it, these both aspects are done. Now, the only thing that you can do is, the temperature and the wind are the only two things that you can branch upon. Since, humidity is also used, left out is temperature and wind. So, let us see what happens in this case. This is another important thing that is sometimes, you may be able to we can calculate it out easily, but I am just going to show you something very quick.

So, temperature has, you have mild and cool are the two values. Temperature has actually 3 earlier you know, that right because if you look into the temperature, the temperature has hot, mild and cool. So, in this case, your temperature is only mild and cool. Here, in data set, temperature has only mild and cool. There are two things, the hot does not even appear here at this point. However, if you look into the wind, what do you get to see? One calm has 3 observations yes and no, how many is and no calm has all 3 as yes no is 0.

Same way, if you look into this gusty has how many observations? 2 observations of this yes equal to 0 no is equal to 2. So, if you do the entropy wind, entropy of what we call as rain and wind, then you can see that, there is perfectly no entropy in this be 0. So, the information gain will be maximum for this and what will be your information gain? So, that will be based on your yes and no in this one. So, you can calculate this out by that process because you have in this one in the rain has 5 observations yes is equal to 1, 2, 3, 3, yes and no is equal to 2. So, if you follow, what we did in the branching node E of bright same way, you can calculate,

$$E(Rain) = \left(\frac{3}{5}\right) log2 \left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) log2 \left(\frac{2}{5}\right) = 0$$

So, at any point of time, one of the quick shortcuts in this algorithm is, if you get any of the entropy value to be 0, then that is a very quick way of recognizing the next independent variable that you can branch upon.



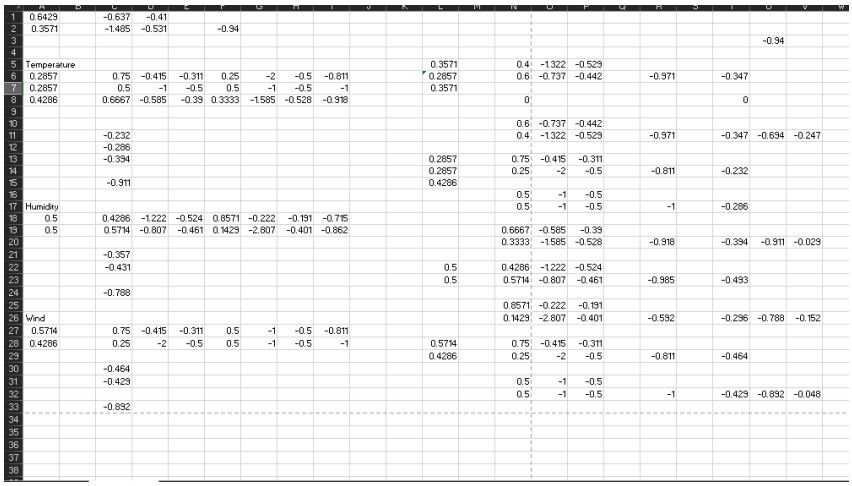
So, if that is the case so here the obvious candidate is wind so the final decision tree. What we can do here is, you can have a weather and weather has 3 branches. The first one is Bright, the next one is Cloudy, last one is Rain and Cloudy. So, if it is cloudy, go for fishing, if it is bright branch on the next independent variable, humidity.

And, if it is there is 2 values to this, there is normal humidity and high humidity. If it is normal, go for fishing it is high humidity, do not go for fishing. And, in the rain, if it comes, it is rain is going on you take the next independent variable, which is wind And, in the wind there is 2 decisions, it is calm or it is gusty. If it is calm, go for fishing, if it is gusty, do not go for fishing. So, now you have a complete Decision Tree with all 5 leaves now.

The final Decision Tree has 5 leaves with definite outcomes. So, the idea is, that you first test the weather, what is the weather like. Now, if a fresh data set of something like cloudy, then high temperature, humidity is also high wind is gusty is given to you the answer is. So, this is Weather, this is Temperature, this is Humidity and this is Wind is given to you the answer is fishing. The question answer will be yes because cloudy, it will just branch on ignore everything and we will just go for yes straight forward.

So, in this whole process you can see. So, somebody gives you another data set, let us say, it is weather comes to be rain and the temperature is high and humidity is normal and wind is gusty. So, then what will happen is, the Decision Tree will say, rain will go into this branch, you look at wind gusty and say no, do not go for fishing, it will ignore the other aspects of it. So, this kind of approach where, from the data set, we have derived rules to decide on Decision Tree, this is the idea of a Decision Tree Algorithm or

Decision Tree Process. I have only explained this using the ID3, there are many other approaches to Gini index, Gart, there are so many algorithms available to this.



1	0.6429		-0.637	-0.41																
2	0.3571		-1.485	-0.531		-0.94														
3																			-0.94	
4																				
5	Temperature										0.3571		0.4	-1.322	-0.529					
6	0.2857		0.75	-0.415	-0.311	0.25	-2	-0.5	-0.811		0.2857		0.6	-0.737	-0.442		-0.971		-0.347	
7	0.2857		0.5	-1	-0.5	0.5	-1	-0.5	-1		0.3571									
8	0.4286		0.6667	-0.585	-0.39	0.3333	-1.585	-0.528	-0.918				0						0	
9																				
10													0.6	-0.737	-0.442					
11			-0.232										0.4	-1.322	-0.529		-0.971		-0.347	-0.694	-0.247
12			-0.286																	
13			-0.394								0.2857		0.75	-0.415	-0.311					
14											0.2857		0.25	-2	-0.5		-0.811		-0.232	
15			-0.911								0.4286									
16													0.5	-1	-0.5					
17	Humidity												0.5	-1	-0.5		-1		-0.286	
18	0.5		0.4286	-1.222	-0.524	0.8571	-0.222	-0.191	-0.715											
19	0.5		0.5714	-0.807	-0.461	0.1429	-2.807	-0.401	-0.862				0.6667	-0.585	-0.39					
20													0.3333	-1.585	-0.528		-0.918		-0.394	-0.911	-0.029
21			-0.357																	
22			-0.431								0.5		0.4286	-1.222	-0.524					
23											0.5		0.5714	-0.807	-0.461		-0.985		-0.493	
24			-0.788																	
25													0.8571	-0.222	-0.191					
26	Wind												0.1429	-2.807	-0.401		-0.592		-0.296	-0.788	-0.152
27	0.5714		0.75	-0.415	-0.311	0.5	-1	-0.5	-0.811											
28	0.4286		0.25	-2	-0.5	0.5	-1	-0.5	-1		0.5714		0.75	-0.415	-0.311					
29											0.4286		0.25	-2	-0.5		-0.811		-0.464	
30			-0.464																	
31			-0.429										0.5	-1	-0.5					
32													0.5	-1	-0.5		-1		-0.429	-0.892	-0.048
33			-0.892																	
34																				
35																				
36																				
37																				
38																				

-Ms. Excel final Demonstration

So, Dr. Prabal Pratap Singh will be showing you some of these standard available algorithms using python and other languages, libraries, where you can do this. But you should understand that, most of the stuff in the python is a black box, you should have a clear idea of how the calculation is done, how each attribute gets selected. So, that whatever the result, that comes out of the black box, you can verify or validate and confirm whether this actually make sense or does not make sense, that is very critical, when you are building a decision Support System.

Because it is the decision maker should ideally know or better know what the black box is throwing at you. So, with this we will conclude the Decision Tree Algorithm, I was hoping to complete this whole thing in an hour, but it taken a longer time, but I hope, you clearly understood the Decision Tree generation through the Iterative Dichotomizer tree algorithm and using the entropy and the information gain matrix to build the decision tree.

So, thank you for your patient hearing and I hope you guys will go through the worked out example and then, using the same thing, you will solve other problems and gain your expertise in this. Thank you very much.