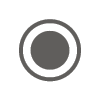
**Meeting with Renusri Ankarapu [Student-PECS]-20241213\_140052-Meeting Recording**

December 13, 2024, 2:00PM

21m 54s

 **Renusri Ankarapu [Student-PECS]** 0:08  
Hello. Hi, good afternoon.  
I'm srik master students from Data science sandwich.  
Placement here.  
I'm here to explain one of my one of the topic from machine learning.  
I'm making as a tutorial.  
Is is a part of my assignment.  
And I I took.  
Topic of my topic is decision tree from machine learning. Decision Tree is basically a supervised learning that is used for classification and regression task by splitting the data set into by splitting the data set into branches based upon based upon their true values and forming like a tree.  
Like a structure.  
In the tree like a structure, it contains node that represents that represents a decision based upon the values.  
And branches.  
Represents the outcome of the decision and the leaf note is a is. Represents the final prediction or outcome of the decision tree.  
And here I want to give an example.  
Of how the decision tree works by sharing my screen.  
OK.  
Let's consider.  
Else if statement considering.  
If.  
Sorry, but my hand hurting.  
If H.  
H equal to less than or equal to.  
18 then.  
And I'm using short terms print.  
College.  
Image better than.  
80.  
Defy.  
It.  
Up.  
Ages less than or sorry.  
Better than.  
Or equal to 18.  
Sorry, less less than or equal to 18 then print college.  
As else if ages, sorry. Else if age is equal to.  
Greater than 18 are greater than or less than equal to 35.  
Then print work and this basically this else if statement can be represented in a decision tree.  
Let's consider a node that is.  
Condition.  
Less than or equal to 18. If the condition is true then it returns yes.  
And it.  
College if the condition is true, then the IT returns college.  
If.  
Mission it is. No.  
It returns.  
Condition is no.  
Then it goes to the next condition that is.  
Better than.  
18 or less than or equal to.  
Age 35.  
This condition is true.  
Then the tree again this in two.  
Then yes, it's true. Then it's.  
Work.  
If it's no, the condition returns no.  
Then it goes again into the subdivisions and this is called root north.  
This this call where the condition returns true is call the leaf node where is the where is. This is our final output of the prediction of the input.  
This is called leaf node.  
And this how basically the decision tree.  
Books inside.  
And let's consider.  
Just consider I'm sorry.  
This concern example.  
Oh, let's consider small example here.  
We we take a feature, one of annual we let's take a feature.  
As an outlook outlook as a feature.  
Basically this is in classification problem.  
Let's see what class, what binary.  
Values we have that is binary values is nothing but yes or no. That is play tennis here.  
Yes, we we basically we classify this data set.  
Or.  
Data set as yes or no as a classification problem.  
That's they.  
They can go out or no on a sunny or rainy day or overcast, depending upon the classified things.  
So we have three different features in outlook that is nothing but.  
And our guest?  
And only starting letters and.  
So in the features in the selected features.  
We have 123-4567.  
So we have yes or no in the play tennis features for the outlook and we got 123-456-7899, yes.  
And we got 12345 finals.  
I.  
Was at 9 years.  
And.  
OK.  
Let's conserve for sunny how many years and how many?  
No, we got.  
Amit.  
Sunny tuna's.  
Good.  
Oh yes.  
Three floors for overcast we got.  
Yes, and 0.  
Nose.  
Rainy we got.  
E.  
Yes, and.  
To loss.  
So.  
Here we got four for the overcast.  
We got the notes.  
Sorry, we got four, yes and 0 notes.  
That means there there is no negative value for it.  
So we got only four years, so it mean this is the final decision, it's.  
So they can. So they go out based upon the values like they got 4 positive values.  
This is a pure data set, sorry pure not.  
Because.  
This is not undergoing further.  
For the for the splitting anymore because they have zero nose.  
If in case we have zero, yes and.  
Four no's, then, they're not going out because we have the negative value for just then this.  
This is this is the final output for this.  
This is the final output and this doesn't go.  
No more split into further splits.  
Whereas here it goes undergoes.  
Splitting based upon the values, because these two are in pure sets.  
Calculate the impurity we use.  
Calculate this impurity and purity we use.  
We use two. We calculate this purity and purities by using two methods, that is entropy and Guinea coefficient or Guinea impurity.  
And entropy.  
Trophy is to calculate the purity of purity is hatches.  
Equal to minus.  
He was positive.  
Log 2.  
He was still minus.  
Eve.  
Ivan.  
To.  
A minus. This is the formula for the calculating the purity of the entropy and.  
On the example we got and the decision tree, that is we got.  
And the data set we took as an example total we got.  
9.  
As we calculated nine years.  
And finals.  
Let us calculate for overcast the purity, the purity of entropy.  
So entropy we got the values.  
This board, yes, and 0.  
Knows so.  
Let us put this upon HS is nothing but the entropy.  
Telling the term explaining.  
And be blessed is nothing but the probability of a outcomes, that is that are positive.  
That is yes.  
Probability of yes is nothing but probability.  
So minus P1P1.  
One lock.  
Sorry.  
It does also let me.  
Sorry, let us consider an example.  
Feature one.  
Say.  
Each one.  
6.  
Yes, and three.  
Knows then it got two different categories, that is.  
See one with values of three.  
Yes, and three.  
Laws.  
And see two category is.  
What 3?  
Yes, and 0.  
So let us calculate the entropy.  
Is.  
S equal to.  
Minus.  
Minus P is nothing but the negative, the.  
Positive and negative probability of yes and probability of no.  
Let us consider probability of three.  
Yes, that's it.  
Probability of yes by total probability of yes is 6 by log of two.  
6.  
Minus probability of no is 0 by 3.  
I.  
Log off two by zero by three.  
Zero by three is is nothing but basically a 0 minus.  
-1 by two log, one by two.  
It's minus.  
So basically a total output is 0 which is called the pure which is the entropy value 0.  
That means it's a pure.  
Entropy value is 0. Then it's called as a purity.  
Your split so it when it when you got the value zero you got nothing to get subdivided and processed for the decision tree and split into different things.  
But when it comes to impurity, then yes, you need it. Undergoes this mode into more steps of splitting the splitting the values.  
Until this process continues in a sequential until until you get the purity split, that is 0 until you get the 0.  
Sure.  
Now also I want.  
I also want to tell.  
Applications that.  
Are used for the application and prod detection, customer segmentation, reductive maintenance or healthcare and also healthcare.  
It got.  
It also got few disadvantages that is prone to overfitting, especially in deep in a deep freeze and sensitive to outliers. When it comes to the outliers, it's difficult and does not autom.  
Atically facilitate.  
Take to educate online.  
So these are few disadvantages that we got in the decision tree and the the.  
The the thing that I explained in this video are few of is our is an example out of my report. I considered a code in my report that.  
I wanted to explain how Decision Tree is work works internally with the basic general examples.  
That's how. That's what I have done in this video and thank you.

 **Renusri Ankarapu [Student-PECS]** stopped transcription