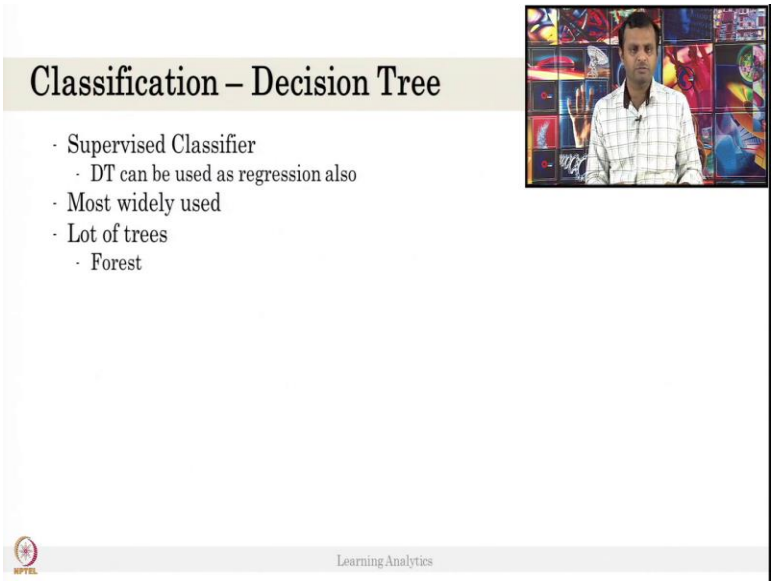
Learning Analytics Tools
Professor Ramkumar Rajendran
Department of Educational Technology
Indian Institute of Technology Bombay
Lecture 9.3
Decision Tree

(Refer Slide Time: 00:28)

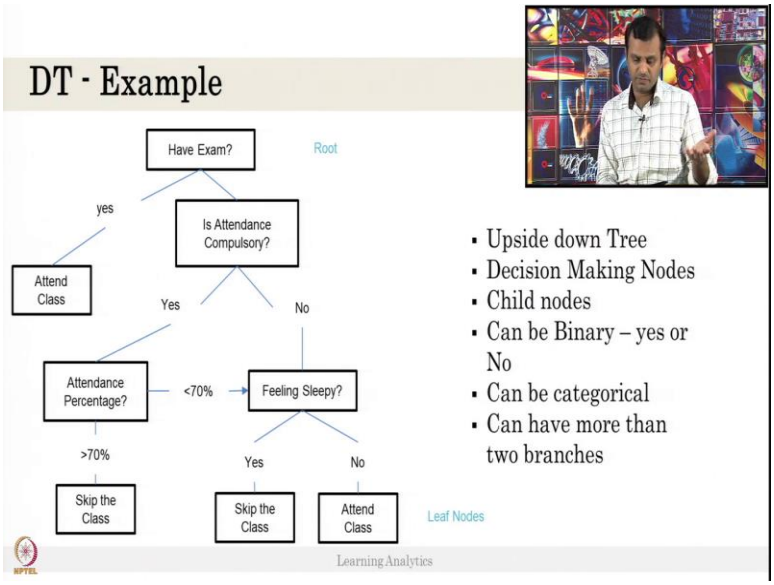


In this video, we will discuss Decision Tree classifier. So, it is also supervised classifiers. Decision tree is also can be used as a Regression that is why it is called CART.

It is very very popular and widely used and most of the researchers report decision tree because it gives a good result. In lecture, we will talk about what the trees simple tree with the root and branches in upside down.

If there are lot of tree instead of one tree, that is I want to construct say 10 trees and pick the, what is the class based on the voting mechanism ensemble classifier, method then it is called Forest. It is called random forest algorithm. So it is good to go for random forest algorithm also . Let us understand what decision tree classifier in this video.

(Refer Slide Time: 01:26)



So, this is a sample decision tree to see, whether student decides to attend the class or want to skip the class. So let us see the two decision, skipping the class or attend the class. If you have exam in that particular class, he will attend the class high probability.

If he has no exam but in this particular course attendance is compulsory, if it is compulsory then the next question comes, if my attendance is criteria then if you have a recquired percentage of attendance like greater than 70 percentage then I can skip the class or otherwise. So there is no exam today and attendanceis compulsory in this course but I have already more than 70 percentage, so I can skip the class or no if attendance is not compulsory it is very simple. So this particular course, I have no exam today no attendance. Now, it is decision of whether you feeling sleepy or not today, if I feel sleepy, I will skip the class, and if I am not feeling sleepy I can go I can attend the class. It is all about the mood of the particular student.

So decision tree classifier is very very famous just because this how, the humans make decision. Suppose before you go to some particular place how you make a decision, you can decide it is important or not, whether what is the cost and all these things you compare. So decision can be made as a tree that is called a decision tree.

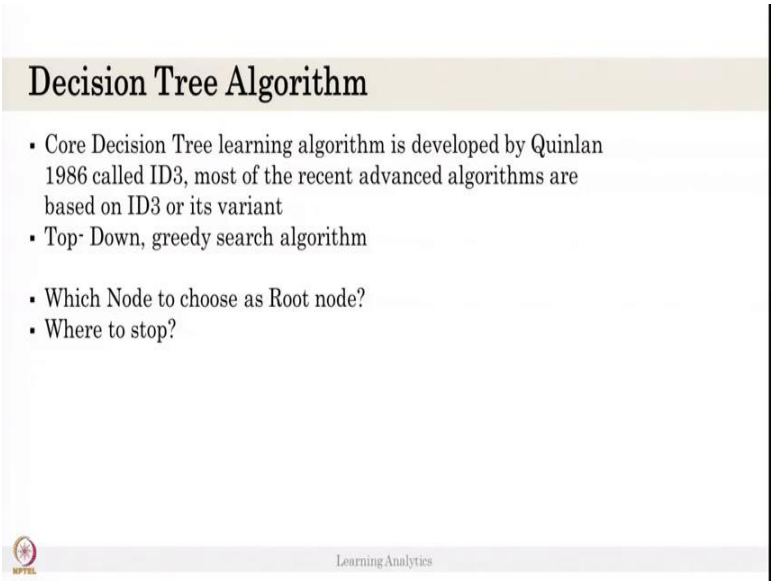
So it very far popular and it is easy to explain to others also. So it is, tree upside down root is in the top and the leaf nodes are all in the bottom. So this is the parent node for this to child's and

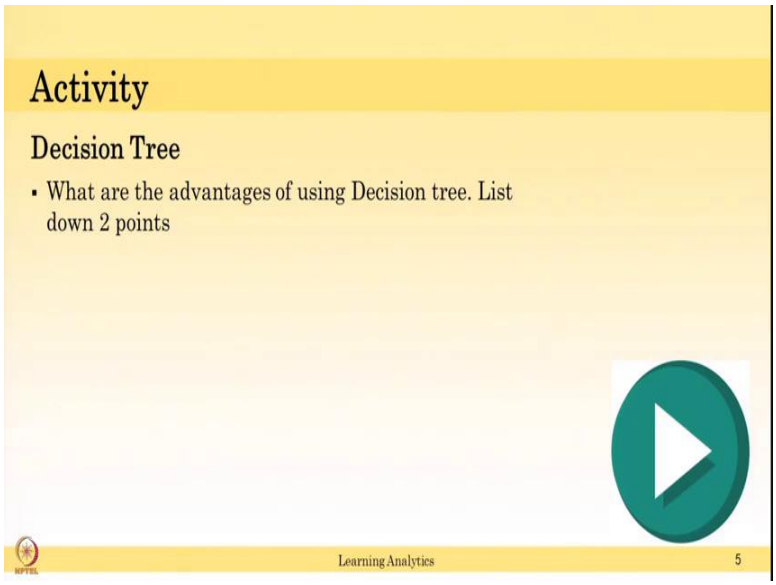
the two child's have a branch two child's and siblings and again, they have leafs, so you have to have a leafs only in the last the layer the top layers are root.

So it can be binary like a yes or no, like skip the class, attend the class or it can be more categorical like what is a percentage marks you get out of. Or it can have a more than two branches.

So, decision tree is upside-down tree and the branch can be binary or non-binary. Also the labesl the variable can be binary or categorical.That is a parent for each child node and root is the root and that is leaves node there is no child for that, that is last node you want to be predict.

(Refer Slide Time: 4:22)



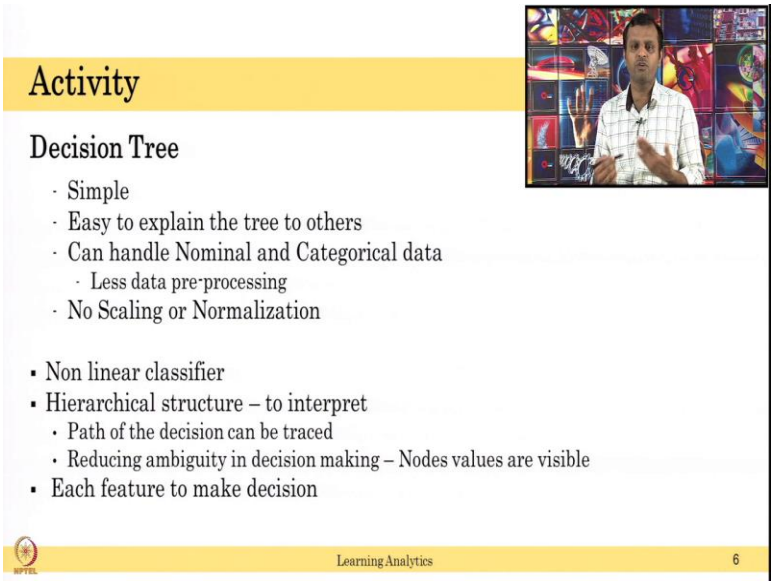


So how do you create a decision tree for a given values or the data you collected from the students interaction with the system something or like that.

So, the decision tree algorithm is developed in 1986 called ID3 algorithm and most recent algorithms like C4 0.5 everything has been developed from the ID3 or the variant of that. So the tree is top down tree and the search algorithm is greedy search will be used, so that is how, you have to search every possible branch if not go to next branch something like that.

So the question here in decision tree which node to choose as a root node. Suppose, there are 4 features, so in a previous example the feature is having exam. How do you choose having exam is the root node? Which node to choose as a root node is very important and also where to stop, that is final leaf node? Can we further continue more conditions and more logics but when to stop it. These two are the key questions in the decision tree algorithm. Let us see that in the next video.

(Refer Slide Time: 05:30)

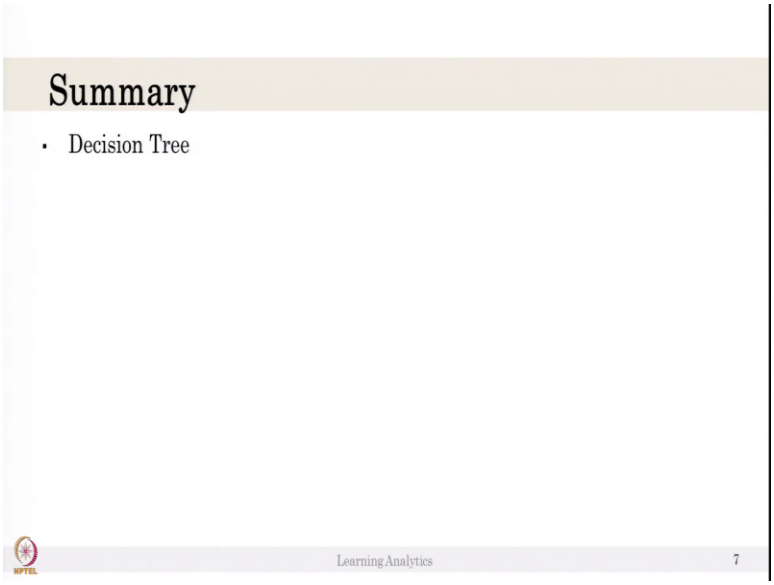


But before that, can you list down the advantages of using decision tree. can you list down one or two advantages compared to other algorithms you are seeing in the this course.

So, decision tree is simple and easy to explain to others because this how humans also make decisio, so we can easily explain to others and they can relate with the daily decision making. Also it can handle both nominal and categorical variable like categorical variable, student pass the exam, meet attendance level's low, high, mid all these things, which means there is less data pre-processing and also no need to scaling or normalizing the values.

It is also a nonlinear classifier. It is not like regression classifier. The structure of the tree is easy to interpret because the path to decision making is traced. So how the particular decision has been made for each thing can be traced from the root to the leave and that reduces the ambiguity in the user when you look at the decision tree when you visually look at the decision tree how the particular decision has been made. So, you can provide informed adaptation recommendation to the students. Also, each feature is considered to make a decision. That is also advantage also disadvantage. Only onefeature is considered at a time other features not considered, but each feature is actually considered to make a decision and that is actually advantage.

(Refer Slide Time: 07:04)



So in this video, we just described what is decision tree and one picture of it and what are the advantages of decision tree in next video we will talk about how to create a decision tree. Thank you.