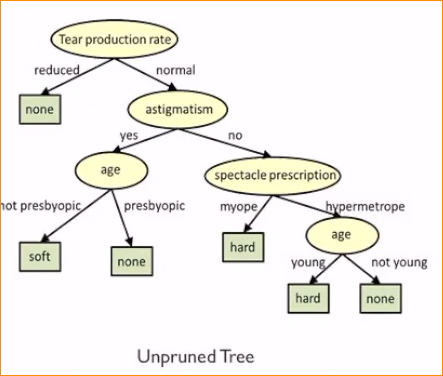
**Decision tree Pruning**



In here,

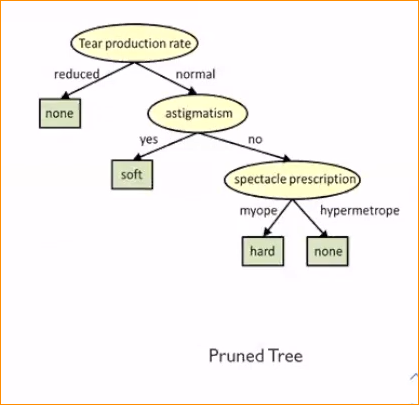
This one have 3 variables – So its like this.

Can you imagine decision tree , that 100 variables???

Some variable is not mush importance when it is predicting your dependent variable.

**That why pruning decision tree is important**

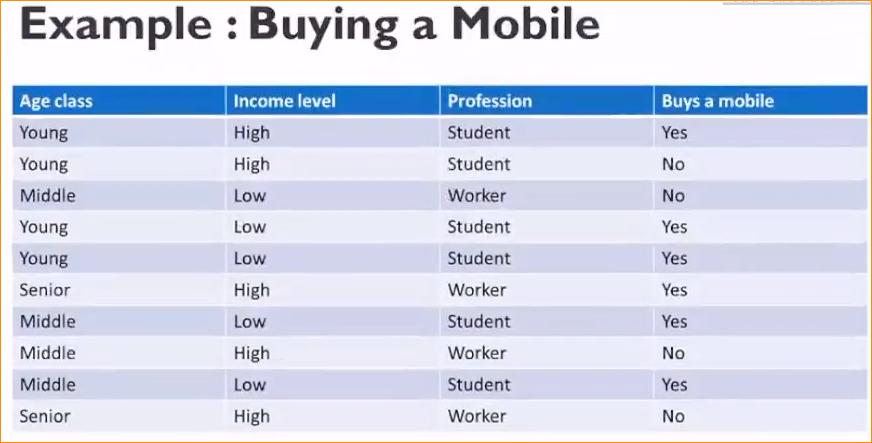
Removing Some insignificant variables, when you predict the categorical variable.

.

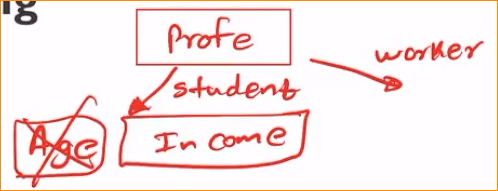
**Decision tree Pruning**

1. Pre-pruning

Stops the non-significant branches from generating. It terminates the generation the generating od new branch based on the given condition.



In here after calculation We choose Profession,



After we have to choose Age or income.

But in the pruning , we consider most important variable.. In here if Age is not consider much than output we just leave it. So we remove unnecessary variable.

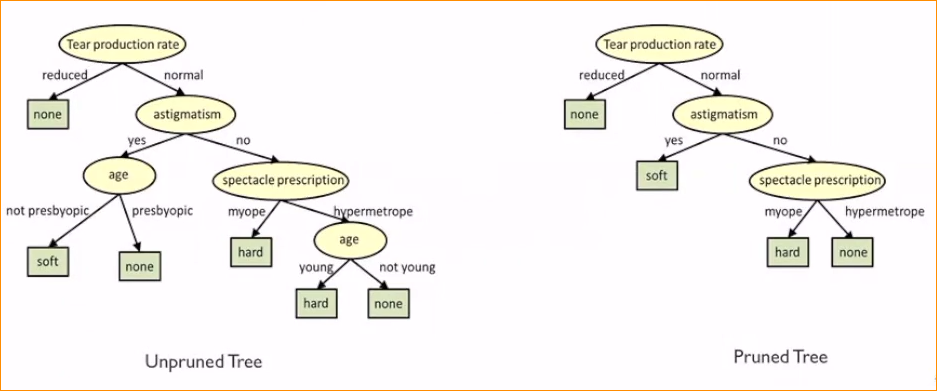
Before making a decision tree.

1. Post-pruning(backward pruning)

Full tree is generated and then the non-significant branches are pruned/removed.

* Cost based pruning
* Reduced error pruning

In here we are creating our decision tree and After checking and filtering.



This provide simple and clean decision tree to you

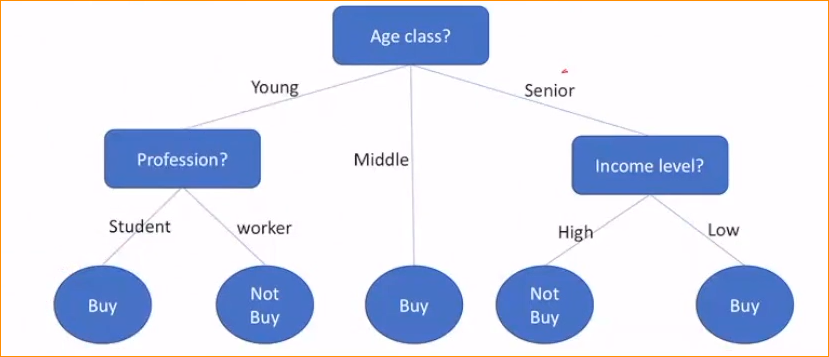
**Rule Extraction from Decision Tree**

This can be done after create our decision tree.

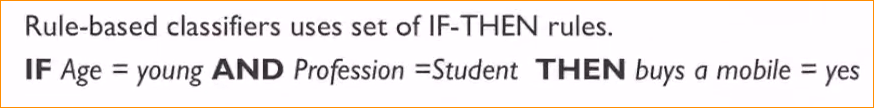
After Extract some rules.

Then use those rules for future prediction or observation.





* If the person is a senior and his income level low – He will buy a mobile phone (Red)
* If the person is Young and he is a student So He will buy a Mobile phone. (Green)
* If the middle age = yes buy a mobile phone.(purple)



**Using these tool making your prediction**

**Model validation –**

Now we know how to train model using decision tree.

Get a mathematical value for the check the **model** check – Working perfect or not.

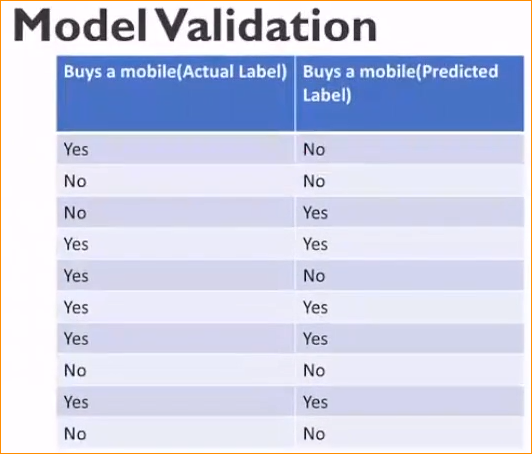
In here we are not going to entire data set to create a decision tree.

In this example we use entire dataset but in huge data set we are using to devide our data set in to two.

Training - 70-80 %

Testing – 20-30 %

So in here we are using our testing data and compare with our predict value and pre labels value.

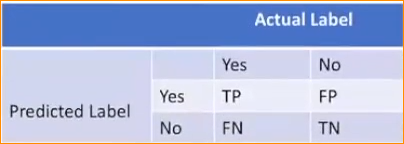


Wrong – Should be “No”

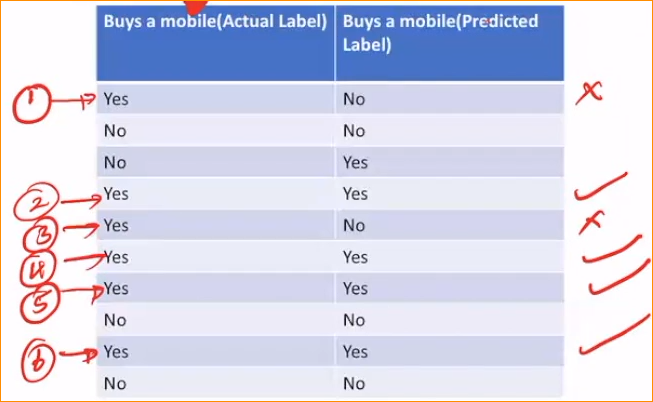
Wrong – Should be “Yes”

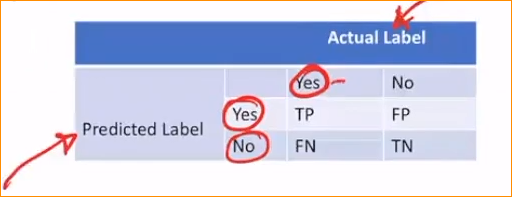
**And making Confusion matrix –**

It is performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicated and actual values.



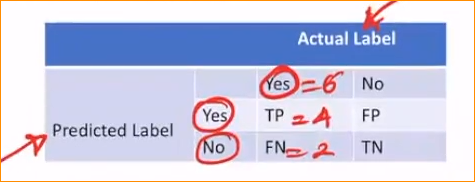
Yes -



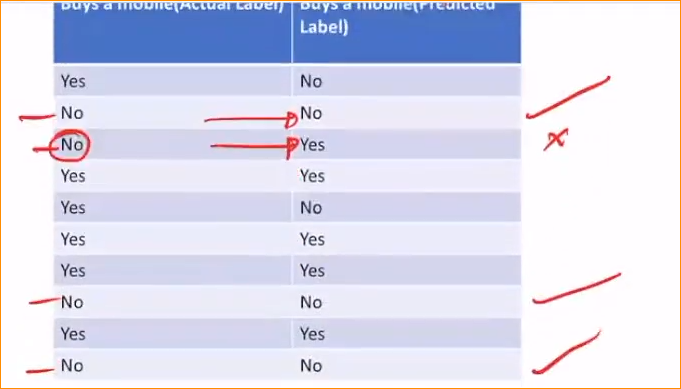


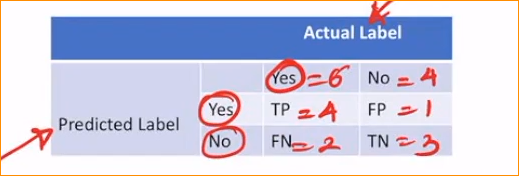
Now take a count –

“Yes” have 6 times but Correct prediction have 4 times 2 wrong.



No –



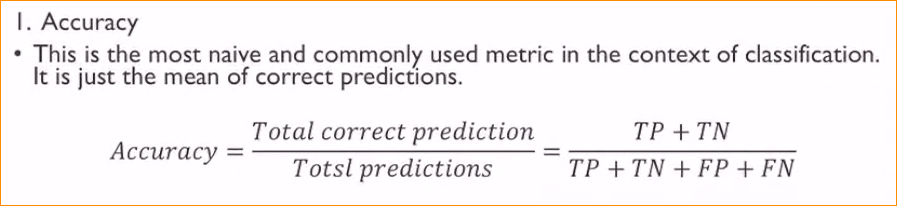


Now we summarize our data correctly or not.

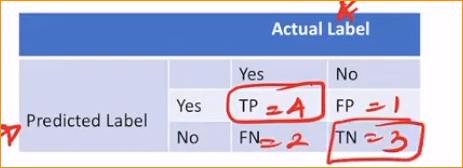
**Performance Evaluation Techniques**

**Accuracy** -

Calculate that you model is correct or not –



Take a correct prediction for up and all prediction for down



Here Accuracy – 7/10

= 0.7

That mean 70 %

Opposite measurement the accuracy call Error = 30 %

Now we can get Idea – Higher classification model in higher or lower accuracy

Other Calculations –

