what is a decision tree?

decision tree help us to take very accurate decision like buying the best bike the available budget prediction of rainfall for the given data sanction of a loan for a person based on his data and also nonlinear regression problems as well as linear regression.

* it is used for decision making
* it is used for prediction purpose
* it is used for regression purpose

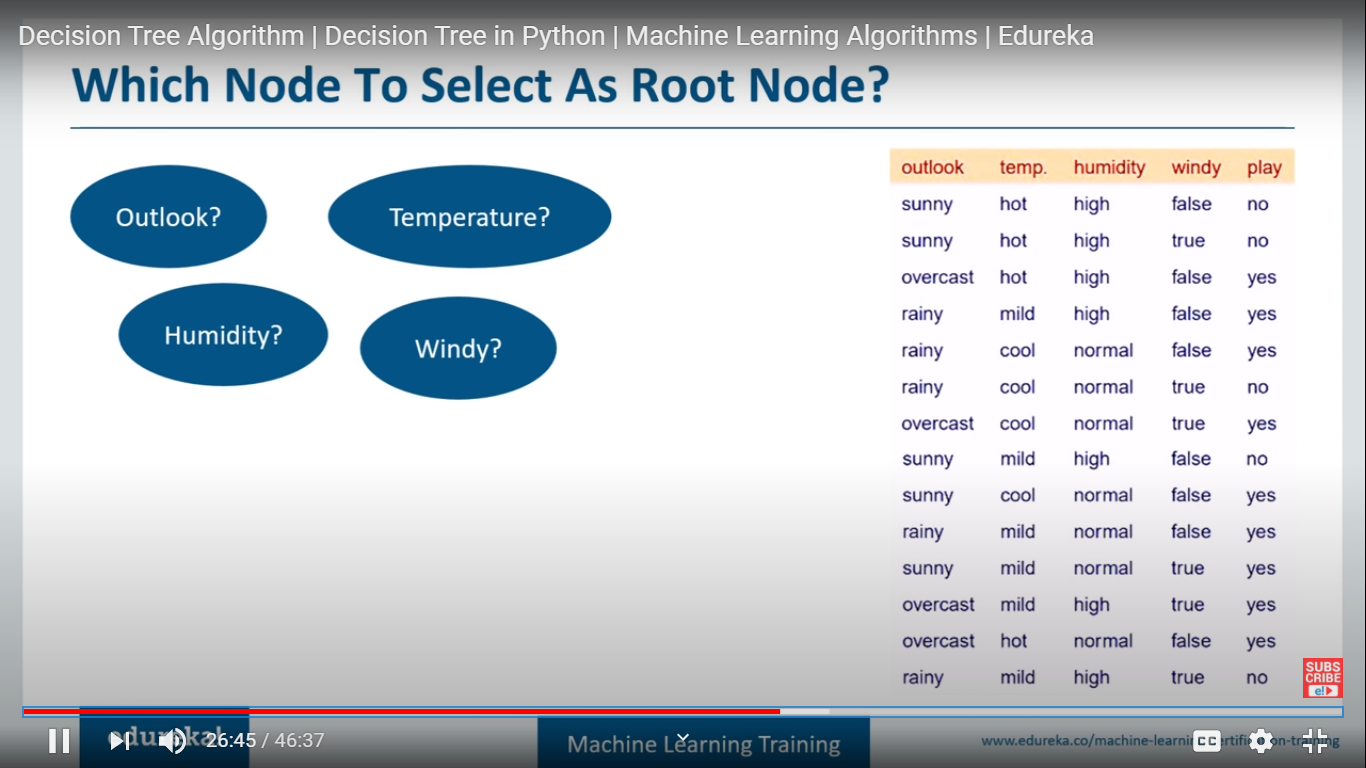
what are the elements of decision tree?

1. root node or parent node
2. child node or branch node
3. leaf node

**parent node** is the first and the main node basis of which the entire tree is created. it is very important select the appropriate parentnode otherwise the length of the tree increases and thereby inefficiency in the tree

**child note** aur branch note are the children of the parent node. you will see in detail in the example.

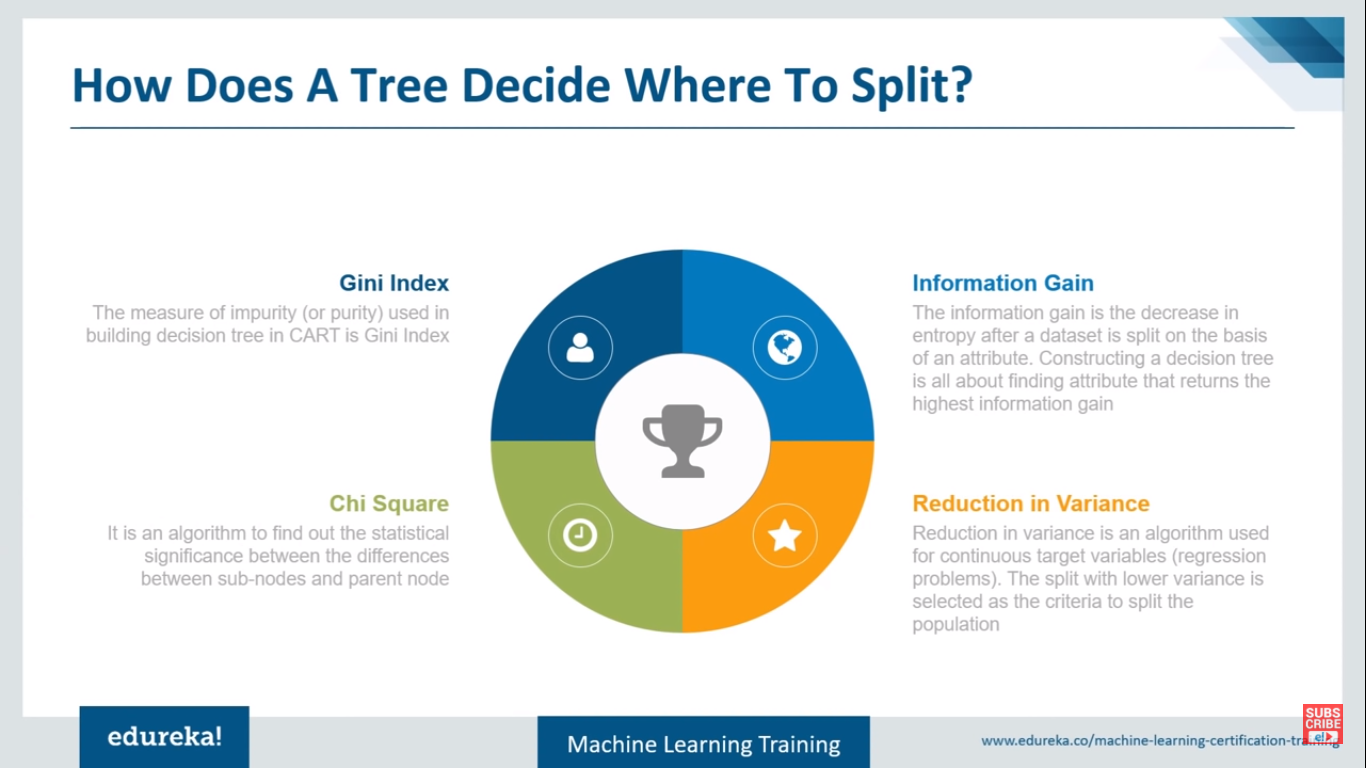
many features are present in the given data. for the decision to give the optimal performance the selection of feature As the parent node plays a significant role.



how to select the parent node?

there are some algorithms which helps us to determine the best feature as the parent node. so what are they?

* Entropy
* Gini index
* Chi square
* reduction in variance.



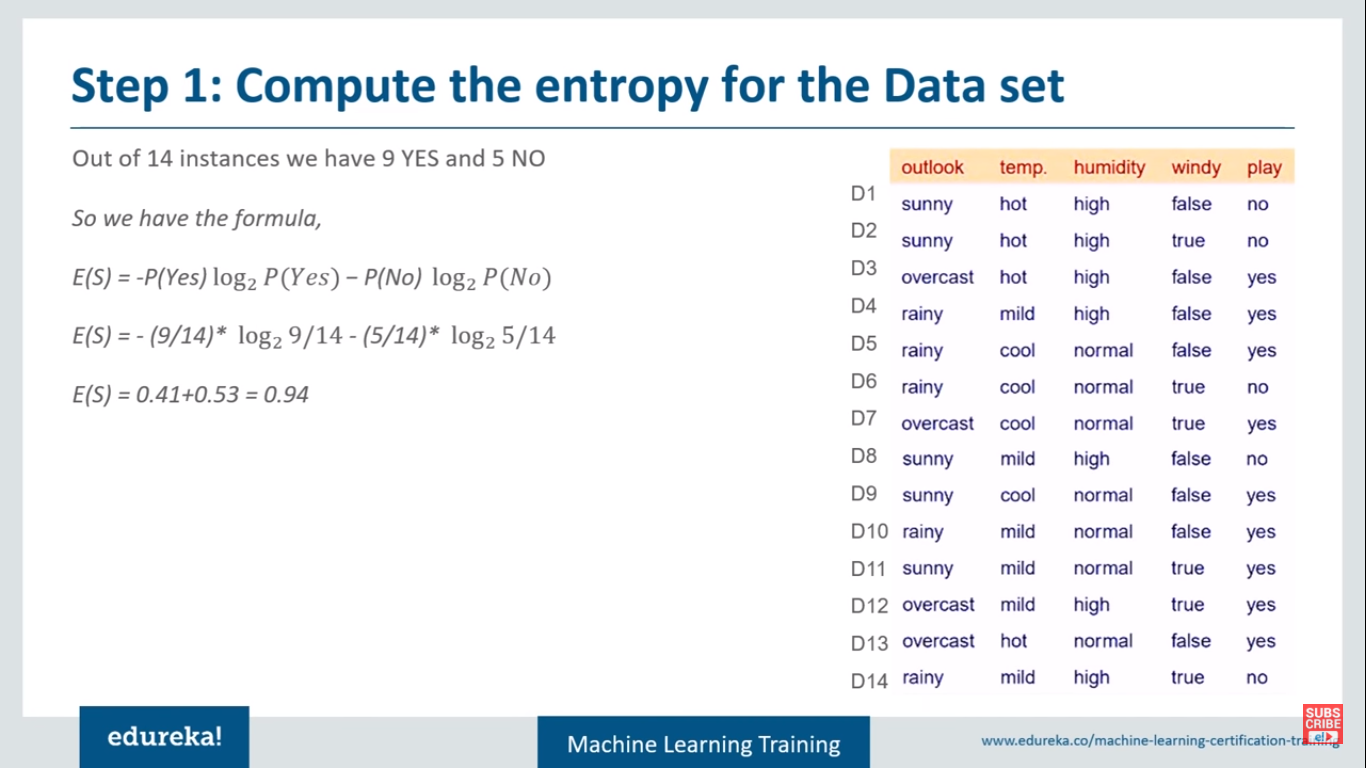
among them first three are used for categorical data and last is used for continuous variable data. let us look at them one by one

**Entropy:** entropy is the measure of randomness. it means in all the available features randomness is present and we calculate that randomness using entropy.

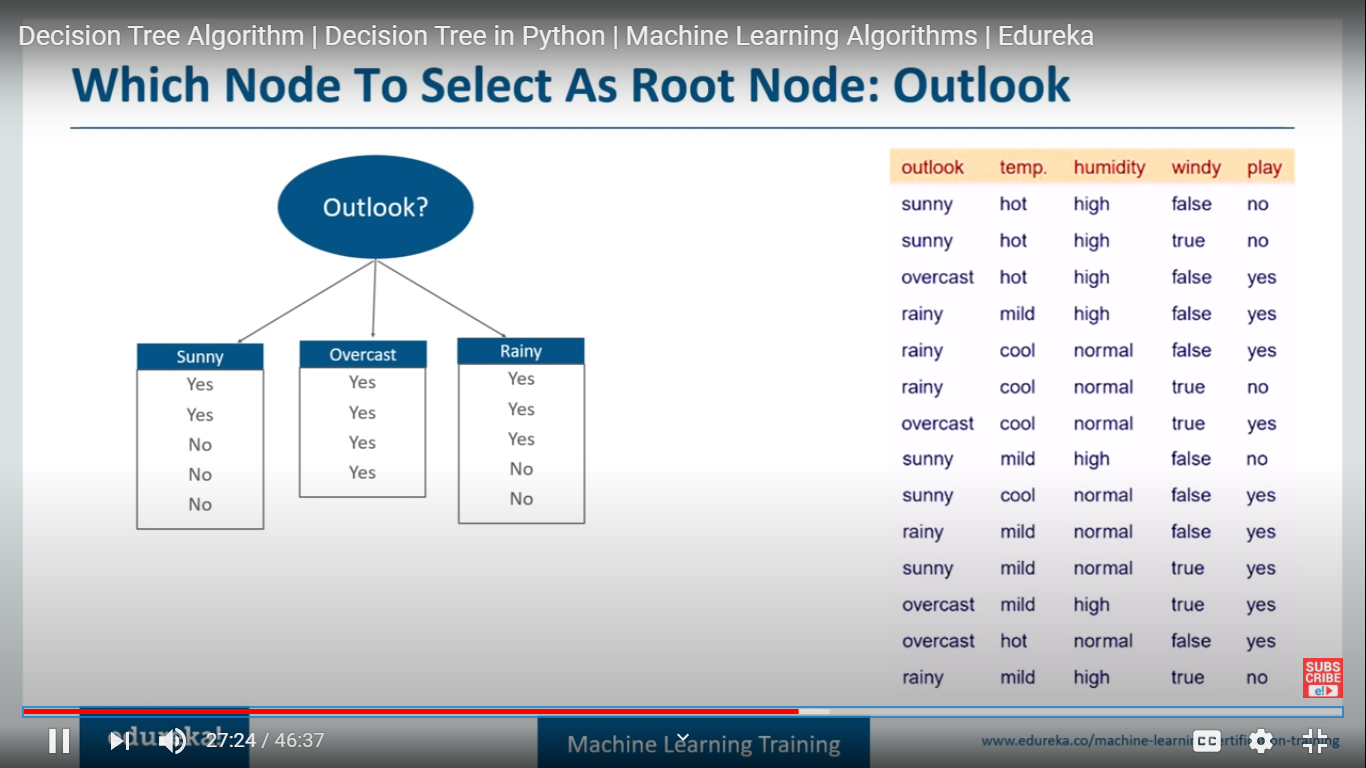
after calculating the entropy we go for information gain. after calculating the information gain the feature which has highest information gain is selected as the root node and the child nodes are also selected with information gain. the one with great information gain are given preference over the others.

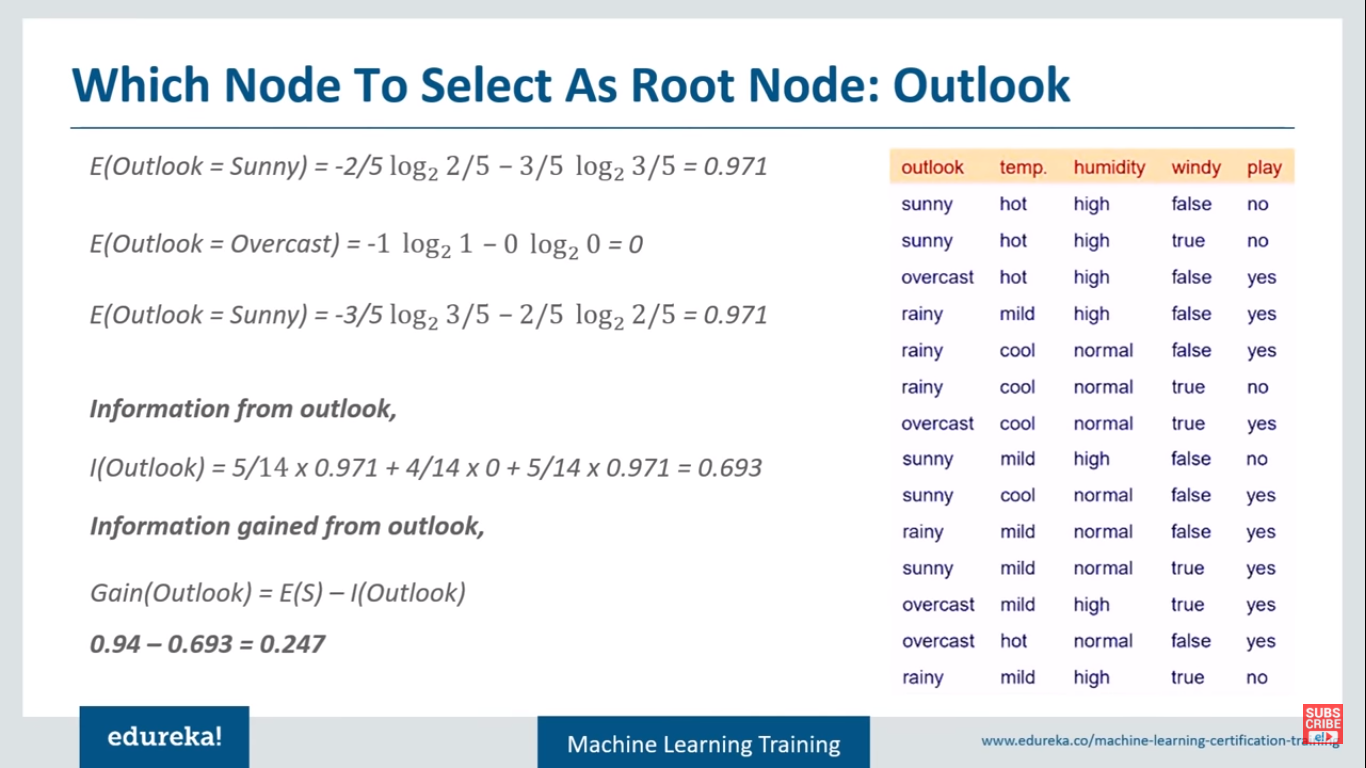
for suppose we have three features F1 F2 F3. among them F3 has highest information gain followed by F1 and F2. so the feature F3 will be selected as the root note by the decision tree.

Here We go with the calculations of entropy information gain using the formulas see the screenshots carefully. just be aware of the content don't remember formulas just be aware of them.

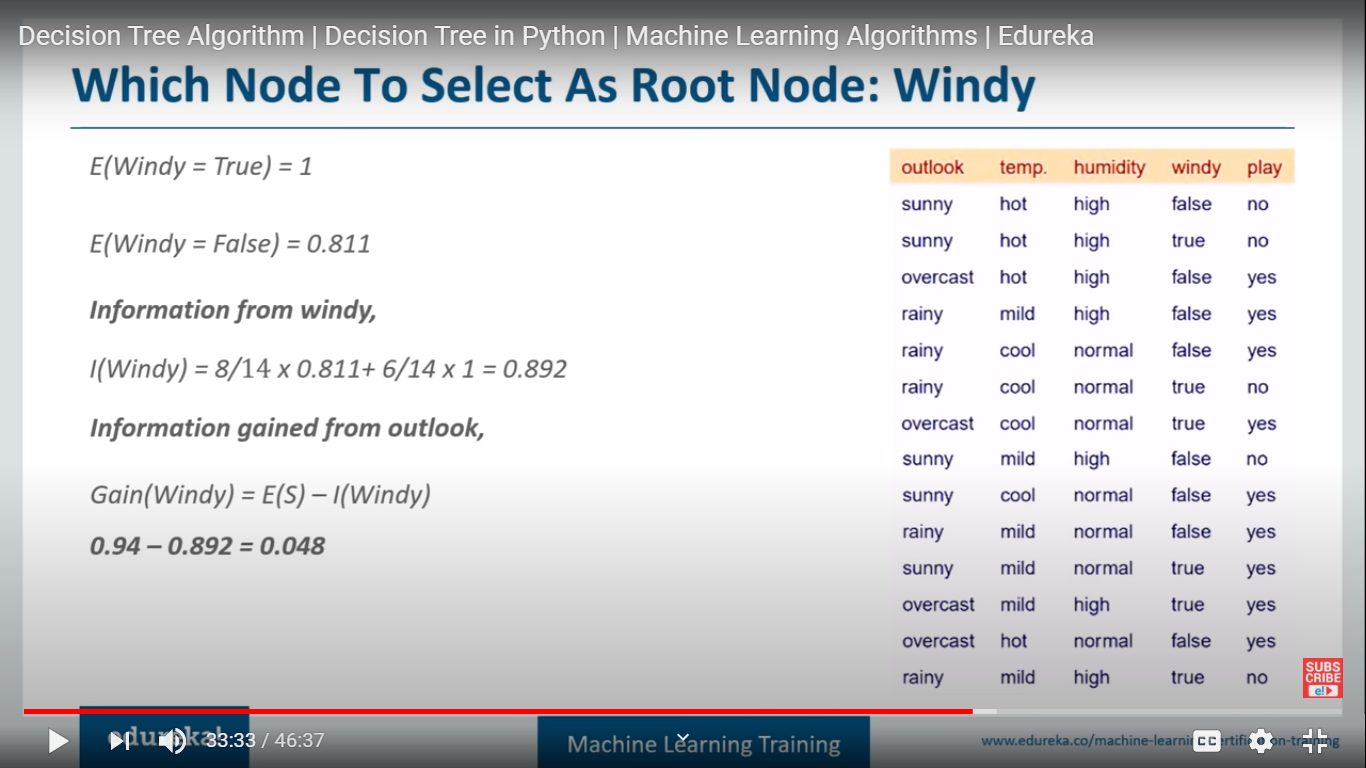


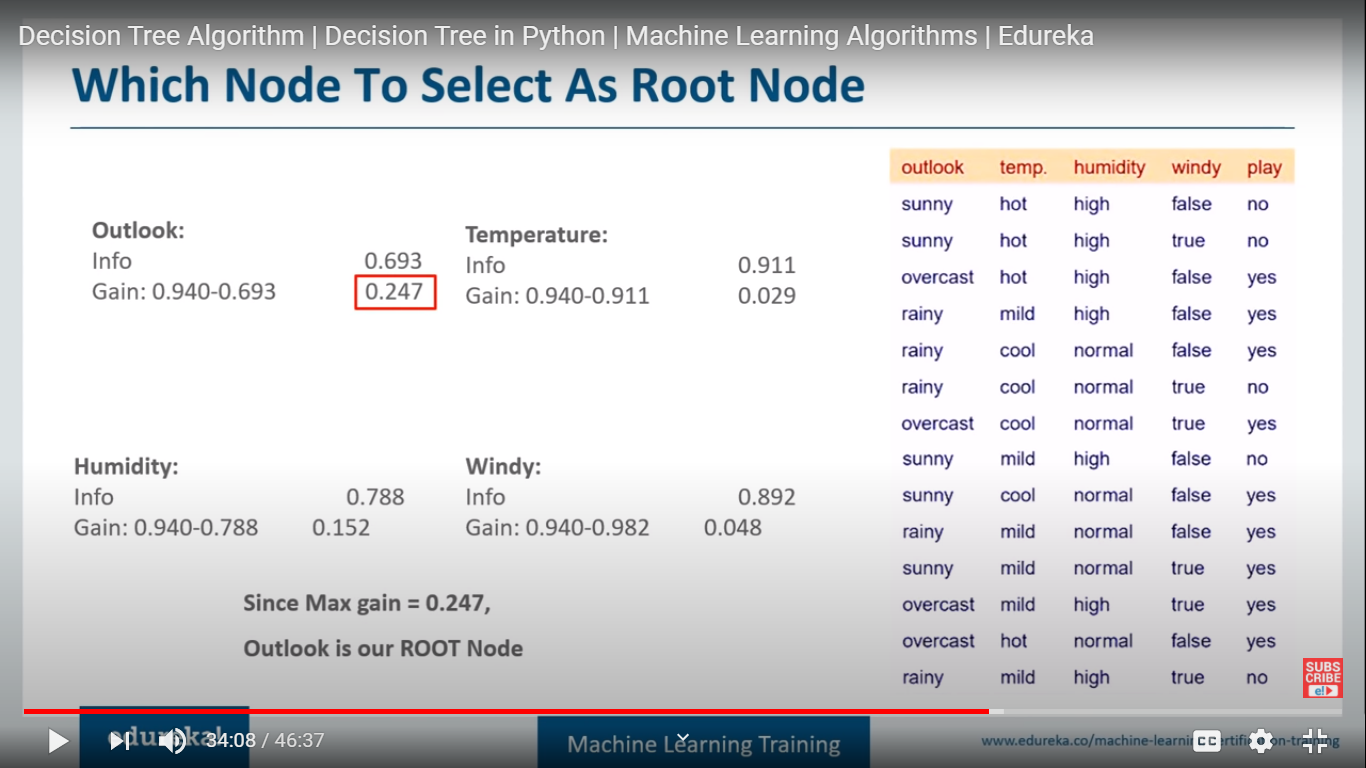
Selecting first outlook and calculating Information Gain





Now We choose windy and calualte the information gain:

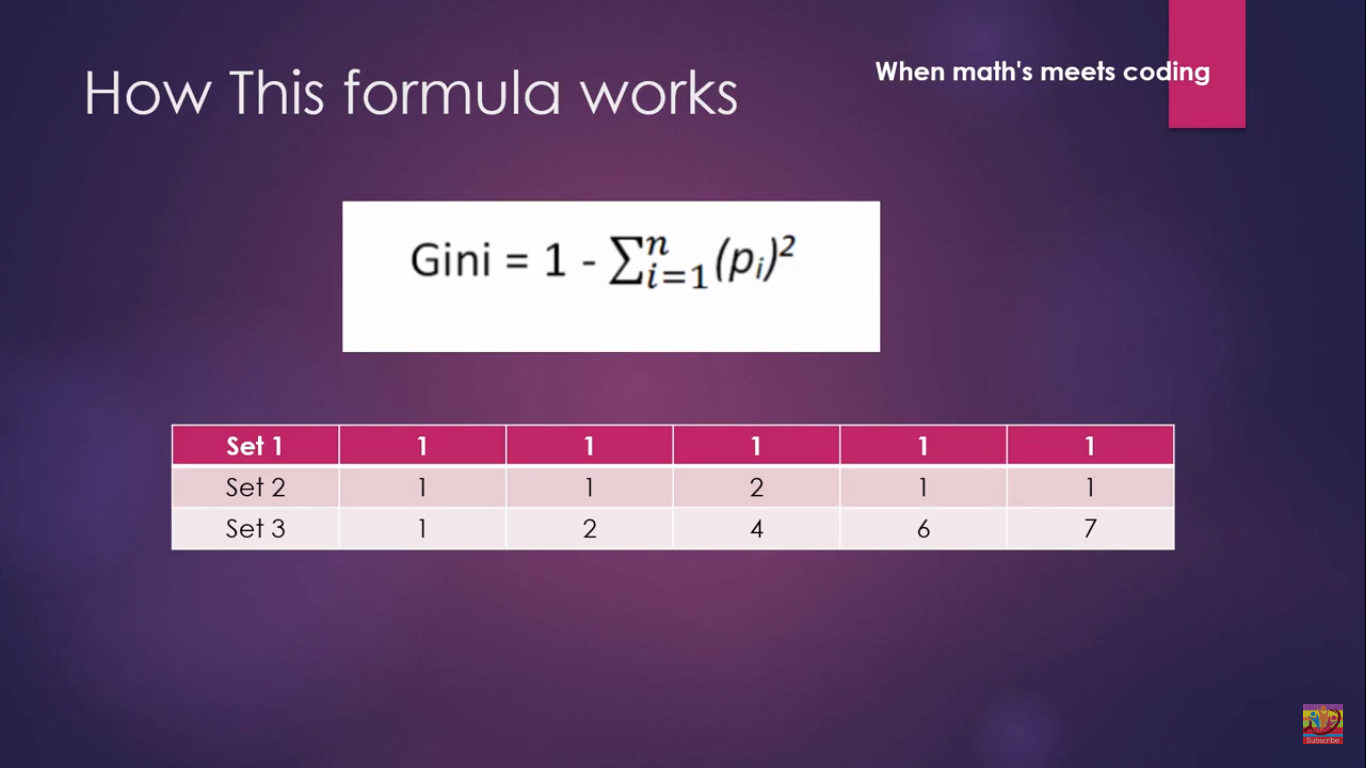


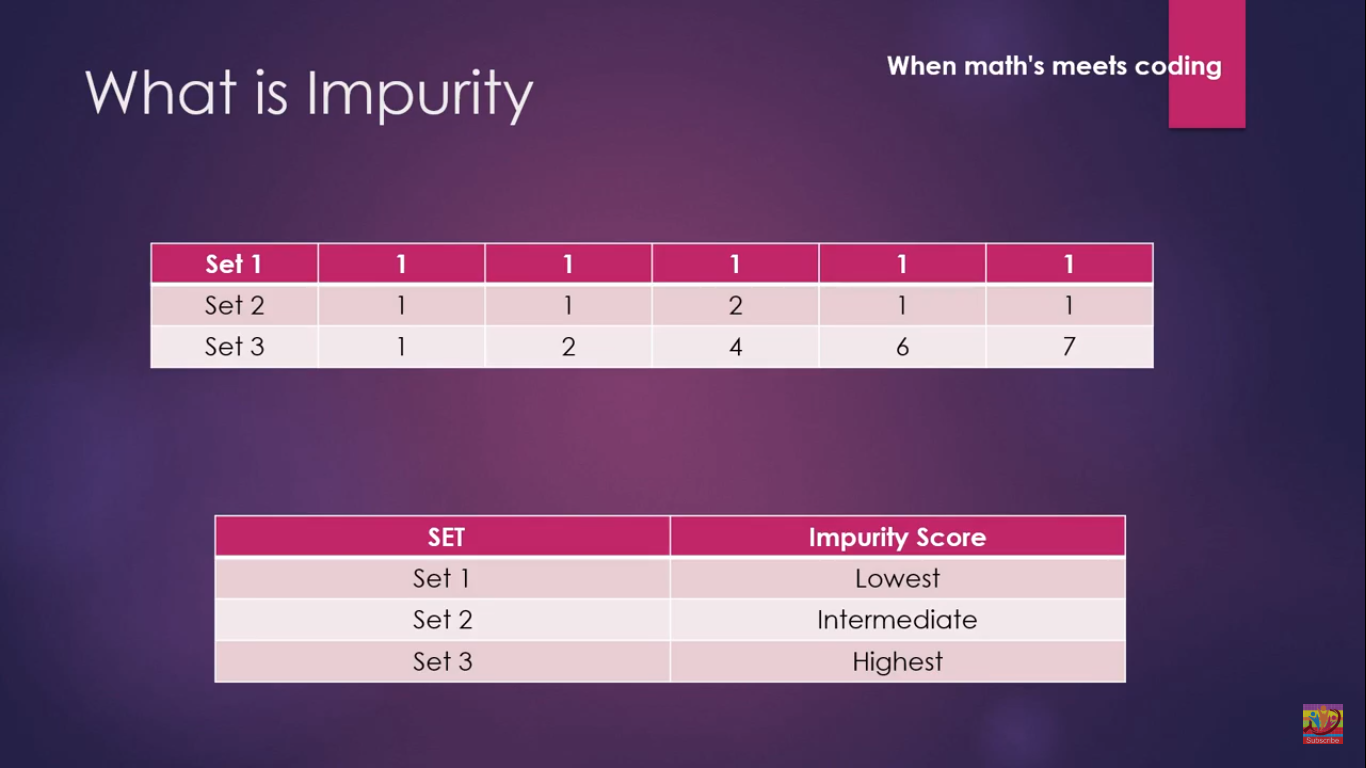


**Gini index**: it checks for the impurity score. the one with the greater impurity selected as the root node.

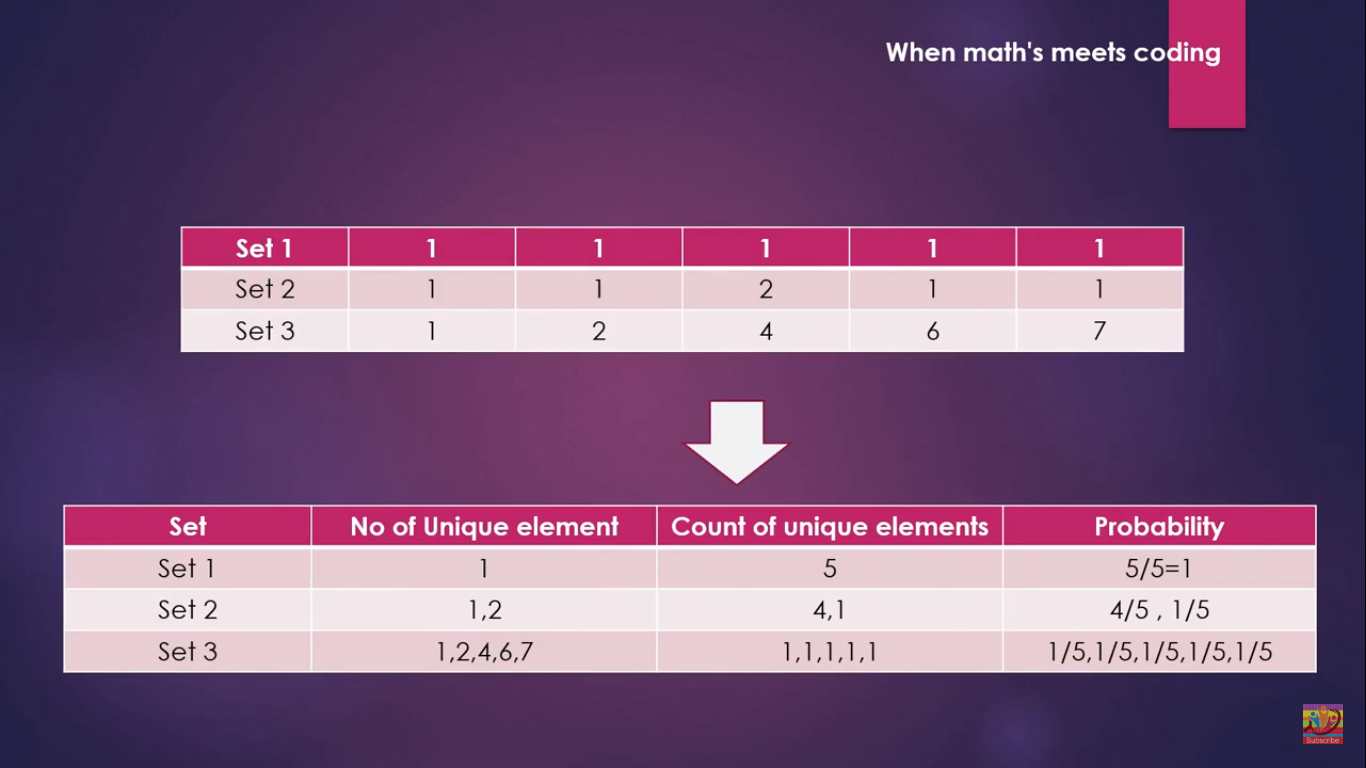
P stands for the probability

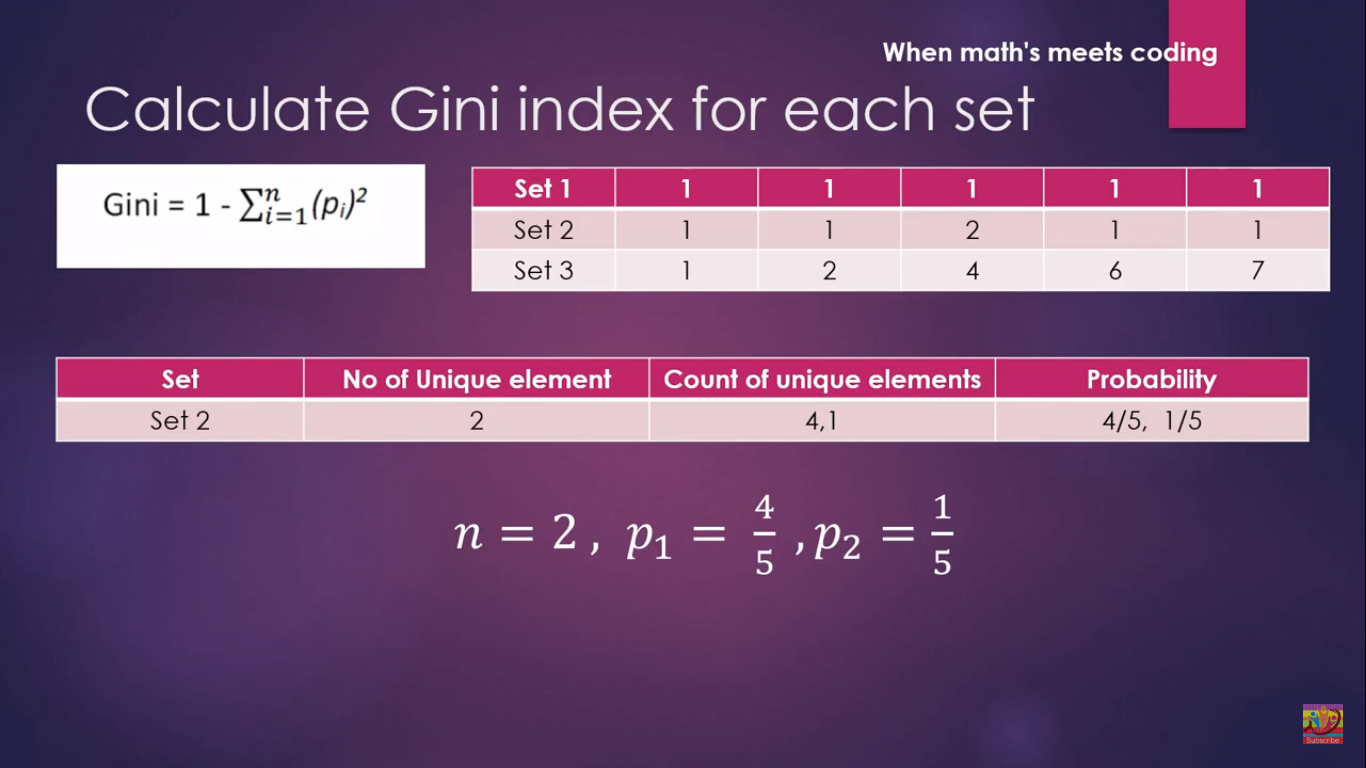
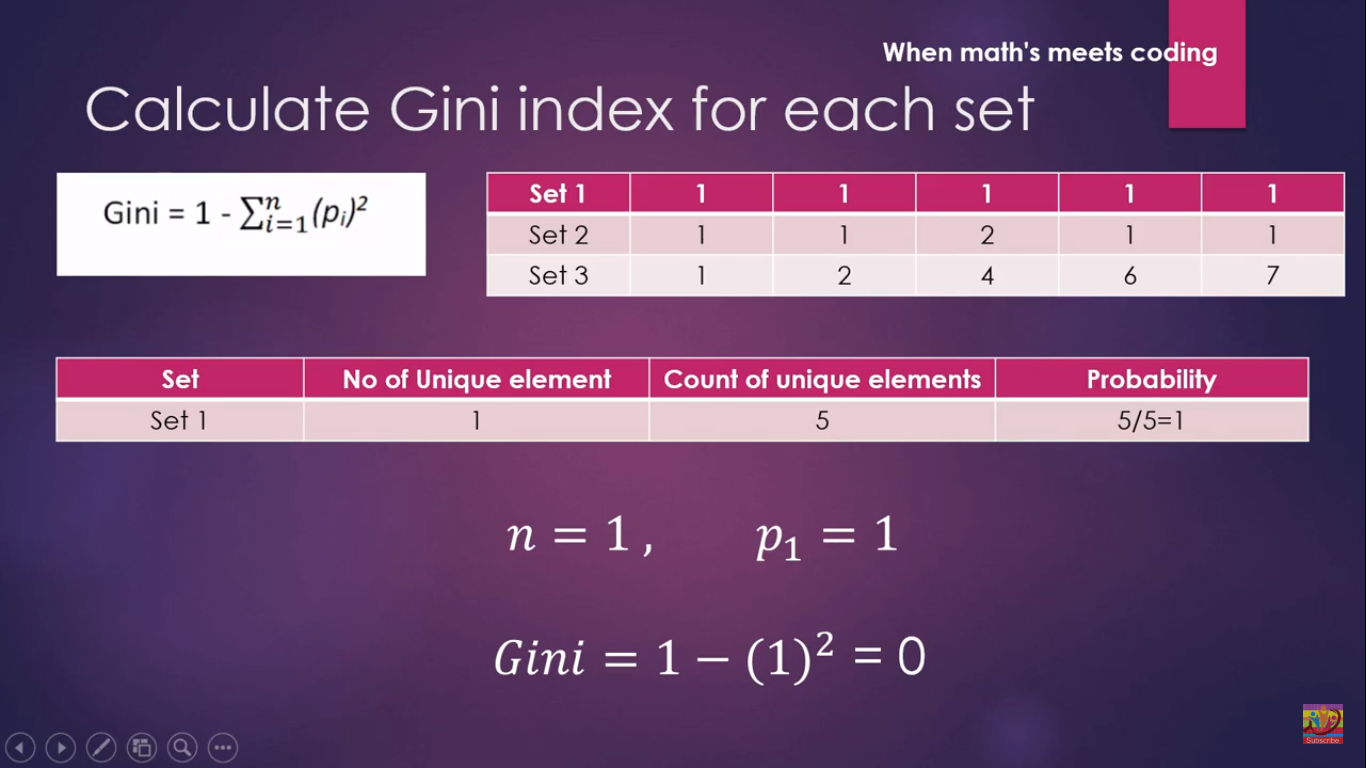
here the n stands for number of unique values. so in a feature greater the number of unique values greater will be randomness and greater the impurities score.



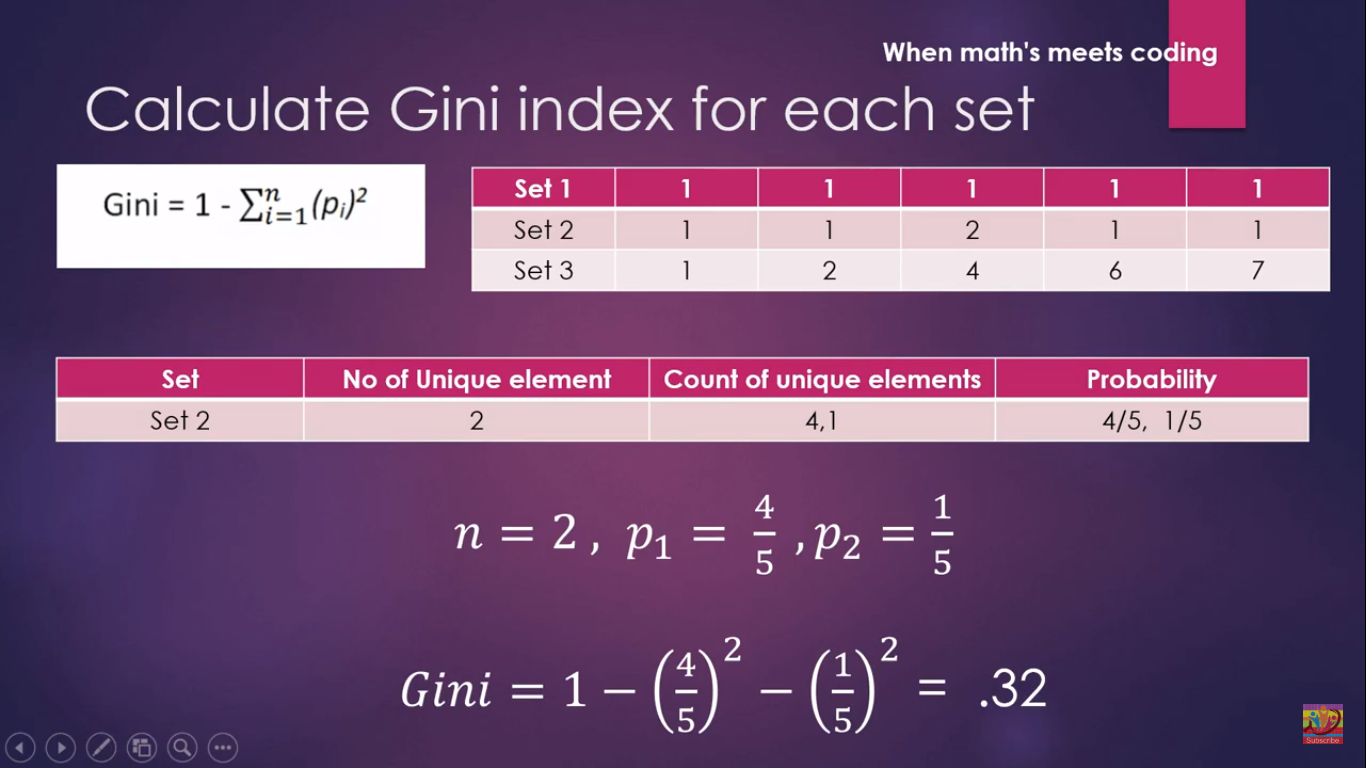
Impurity score depends on the no of the unique values present in the given feature. As set 3 have more unique values its impurity is highest.

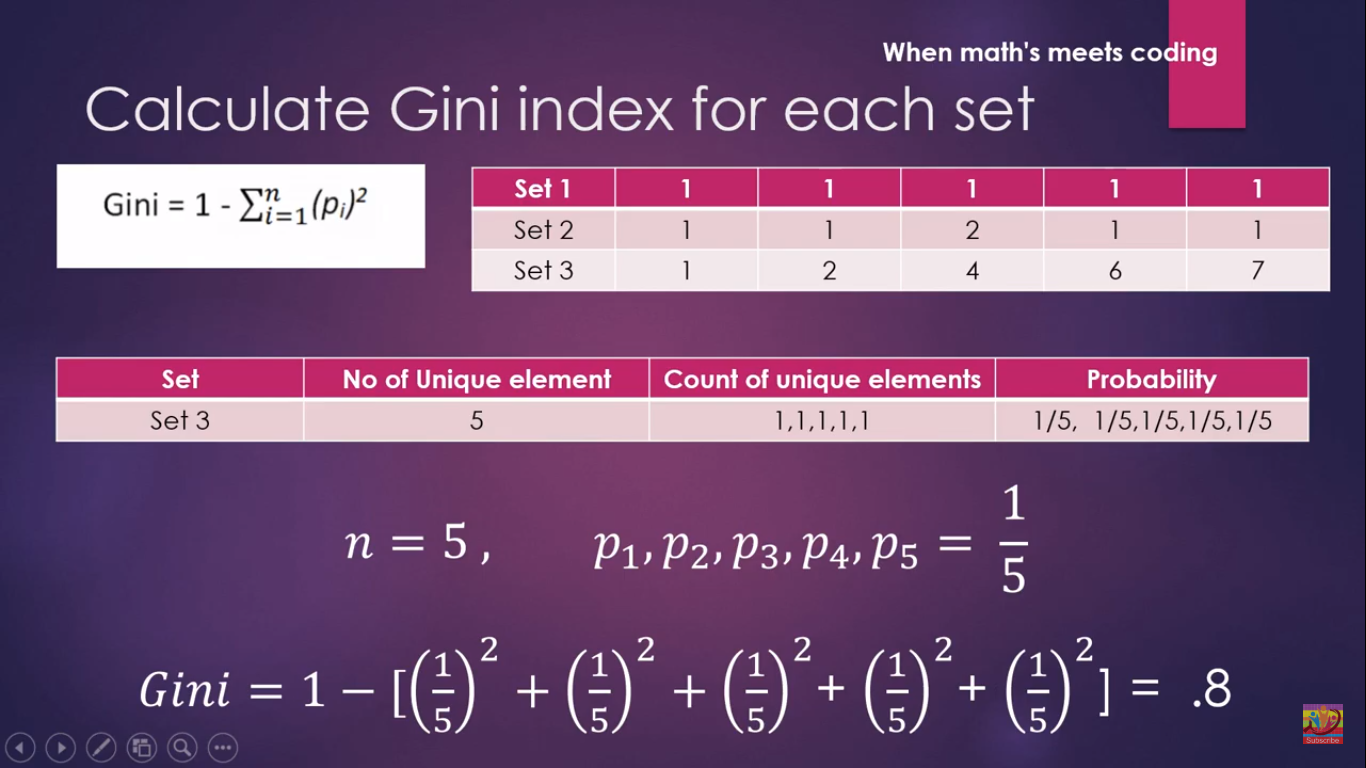
Now we calculte the probablity of the each unique value present in the sets.



Set1 Gini Index:****

Set 2 Gini Index:

****

Set3 Gini Index:****

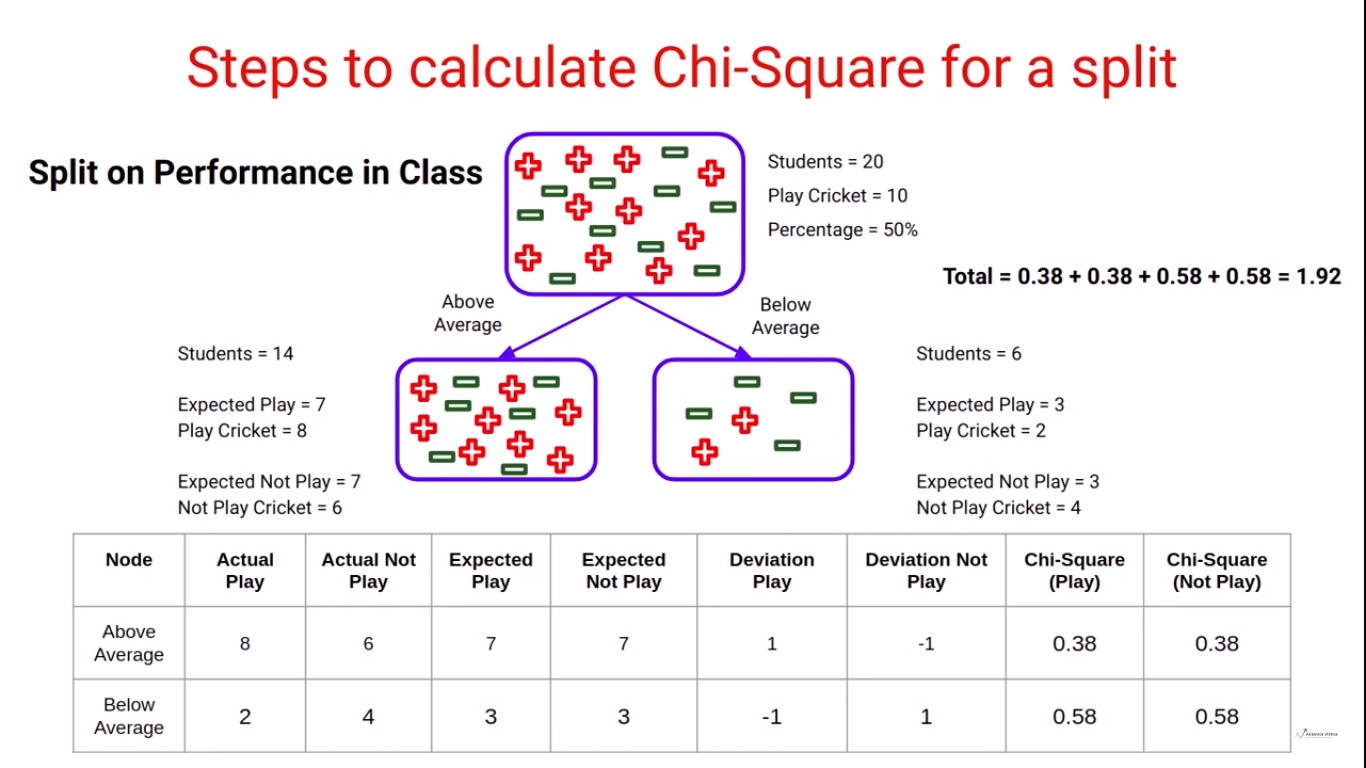
As we see the set3 is having the gini index more so it is selected as the root node.

if a feature having all the elements same then the feature will have zero Gini index as the Gini index zero then it is eliminated for being root node.

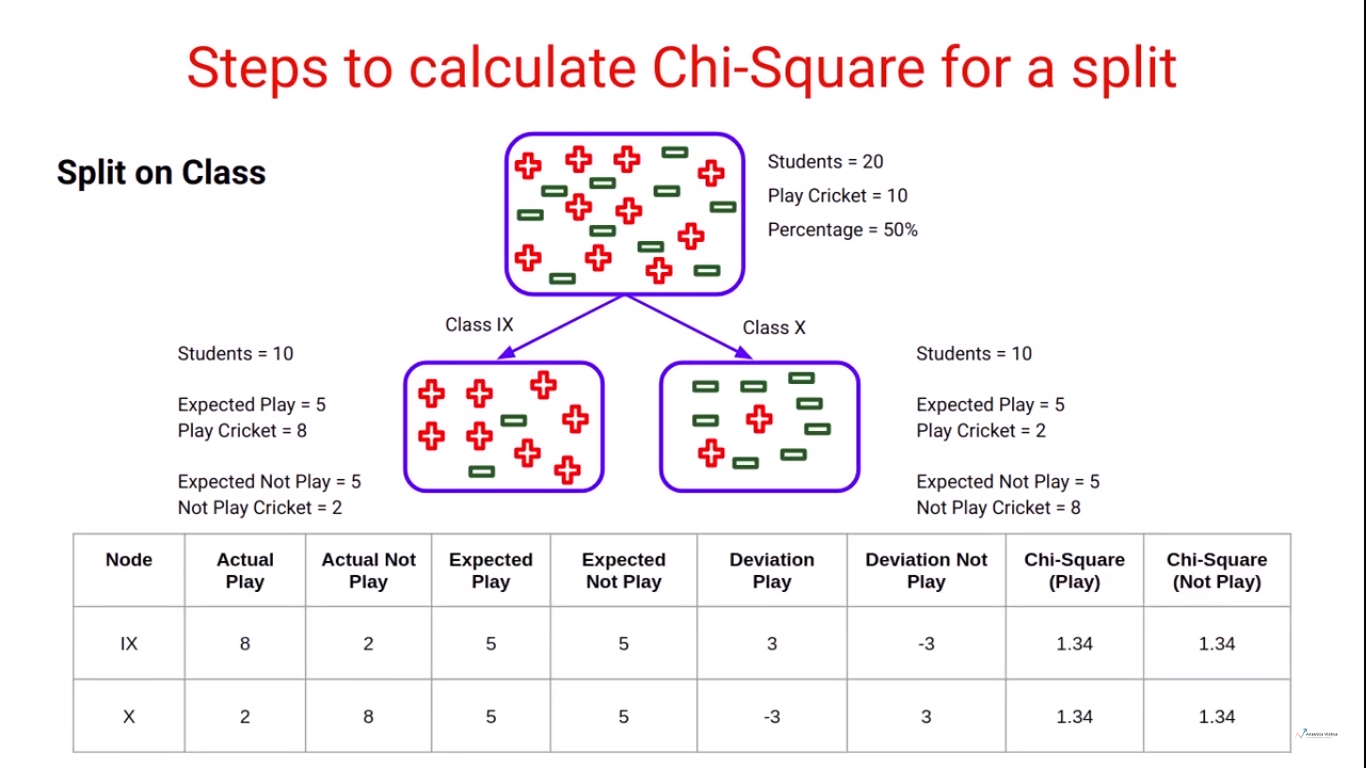
**Chi square**: I said before Chi square deal with categorical data. it is somewhat different fromthe above measures. it select the feature with greatest Chi square value.

group of 20 students 10 bplay the cricket while 10 don't play. it means 50% of students play cricket and 50% don't. here students who play cricket and don't can be predicted using two features percentage and class

when used percentage as feature. 14 students fall under above average and six students below average. as we see 50% play and 50% don't. the algorithm thinks 7 students play 7 students don't above average and 3 students play 3 Students don't from below average. But in actual they differ. we calculate chiSquare play and Chi square don't play and add them.see the screenshot below



like the same making class as the feature we do the same.



Comparing the Chi-Squre Values:



as we see the feature having the greatest Chi square value is selected as the root note.