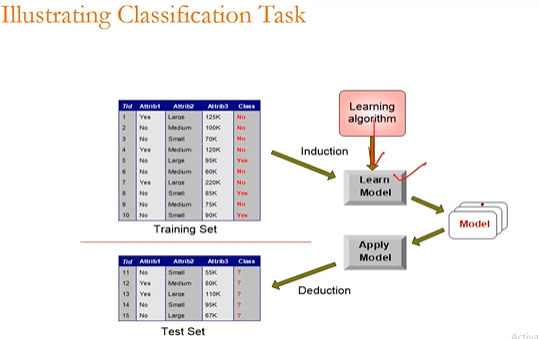
Classification Steps:

Given a collection of records (Training Sets)

Task: Find a model for class attribute as a function of the values of other attributes

Goals: Previously unseen records should be assigned a class as accurately as possible.



Example of Classification Tasks:

Classifying credit card transactions as legitimate or fraudulent

Categorizing new stories as finance, entertainment, weather

Classification Techniques:

Decision Tree based Methods

Neural Networks

Naïve Bayes

Support Vector Machines

K- Nearest Neighbours

Divide and Conquer

Decision trees are built using a heuristic called recursive partitioning. This approach is commonly known as Divide and Conquer. Coz it splits data into into subsets, which are then split repeatedly into even smaller subsets.

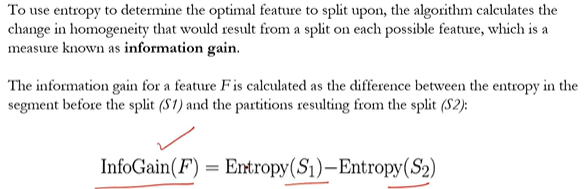
Entropy, a concept borrowed from information theory that quantifies the randomness or disorder within a set of class values

Data Sets with high entropy are very diverse and provide little information about other items that may also belongs to that data set.

The Decision tree hopes to find splits that reduce entropy, ultimately increasing homogeneity within the groups.

If Variance is zero for a column (All values are same means variance is 0)

then that column is not useful.



Entroty (S) = 

