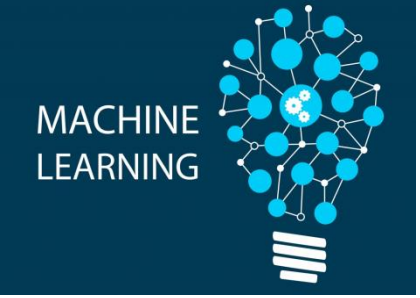
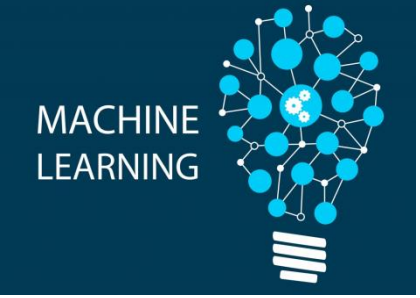



# Decision Tree – Classification

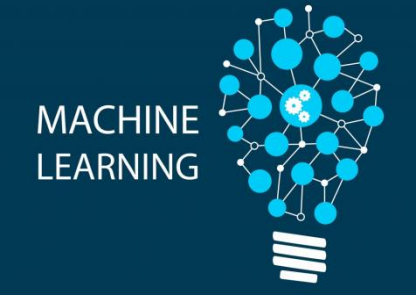
Diptangshu Banik

Twitter - @dipbanik




# Overview

- Decision trees are of two types – Regression and classification
- **For classification, It creates leaves or pockets so that the maximum number of data points of the same class lie in that split.**
- If an input value falls in a leaf, the prediction for that input will be the mode of the classes of the data points in that leaf.
- How and where these splits are conducted are determined by the algorithm.




- Split occurs on the basis of information entropy(is the split increasing the amount of information that we have about our points). The split tries to minimize entropy to achieve that.

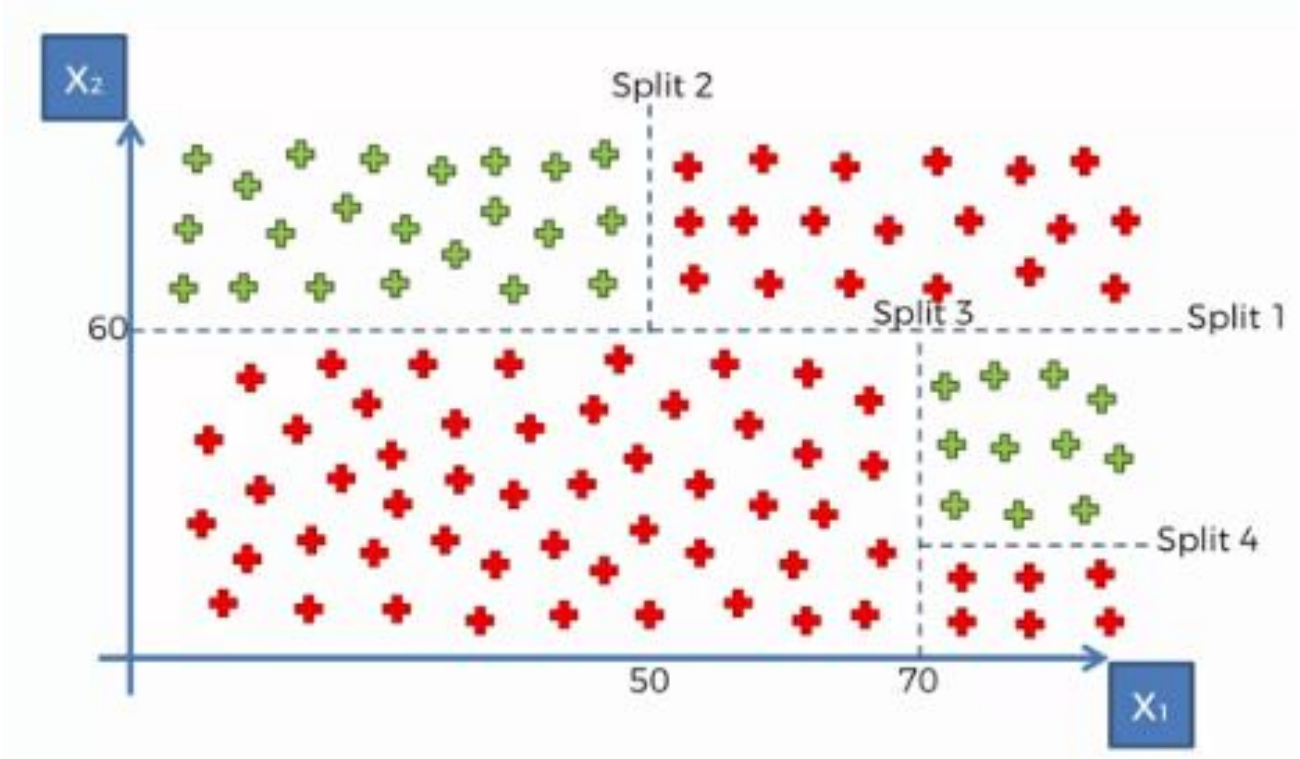
- The split stops when we cannot add any more information about the leaf or the new leaf has less than say 5% of total points of the actual leaf.

- Instead of looking at all the points in our dataset, we are looking at specific cluster which localizes the prediction and can be more accurate.

- If the data is non linear, this is a good option.

MACHINE LEARNING

- Our data set -



MACHINE LEARNING

- Tree -