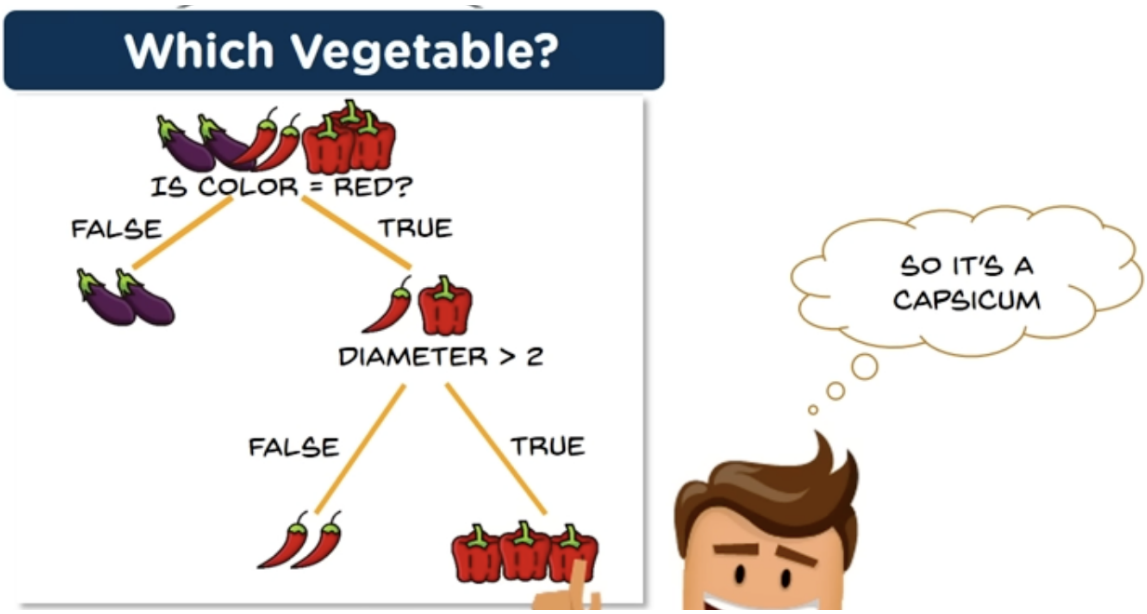
Decision Tree

Decision Tree In Machine Learning | Decision Tree Algorithm In Python |Machine Learning |Simplilearn

🔥Professional Certificate Course In AI And Machine Learning by IIT Kanpur (India Only): https://www.simplilearn.com/iitk-professional-certificate-course-ai-machine-learning?utm_campaign=23AugustTubebuddyExpPCPAIandML&utm_medium=DescriptionFF&utm_source=youtube

https://www.youtube.com/watch?v=RmajweUFKvM

Decision Tree is a tree shaped diagram used to determine a course of action with each branch of the tree representing a possible decision, occurence or reaction.



Decision Trees can be used in Classification and Regression :

- In classification : the tree will determine a set of logistical if then conditions to classify problems.
- In regression : when a target variable is numerical or continuous.

Advantages of using Decision Trees :

- Easy to understand
- Little requirements for data preparation
- Can handle both numerical and string data
- Non linear relationships doesn't affect the performance

Disadvantages of Decision Trees :

- Overfitting : when the algorithm learns the database too well and performs poorly with unseen data because it captures noise.

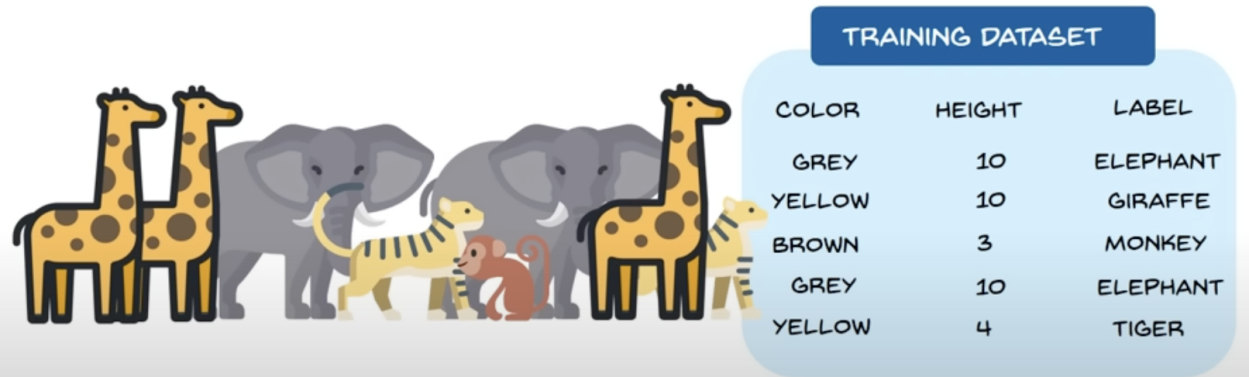
- Model can be unstable due to small variation of data

Important terms :

- Entropy : measure of randomness or unpredictability in the dataset.
- Information gain : measure of decrease in entropy after the dataset is split.
- Leaf node : carries the classification or the decision, final node at the bottom of the tree.
- Root node : top most decision node

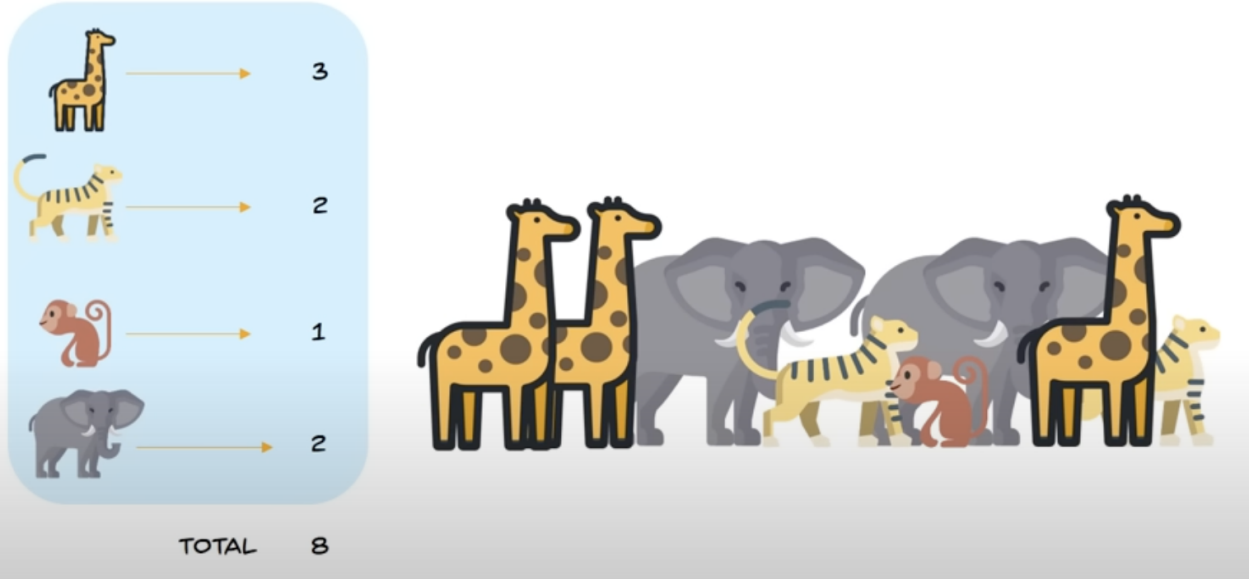
How does a decision tree work :

Consider we have a database with high entropy :



- Frame the conditions that splits the data

How to calculate entropy :



LET'S USE THE FORMULA

$$\sum_{i=1}^{k} P(valuei).\,log_2(P(value_i)$$

ENTROPY = $(\frac{3}{8})\ log_2(\frac{3}{8}) + (\frac{2}{8})\ log_2(\frac{2}{8}) + (\frac{1}{8})\ log_2(\frac{1}{8}) + (\frac{2}{8})\ log_2(\frac{2}{8})$

ENTROPY=0.571

The program will calculate the entropy of the dataset after every split to calculate the information gain.