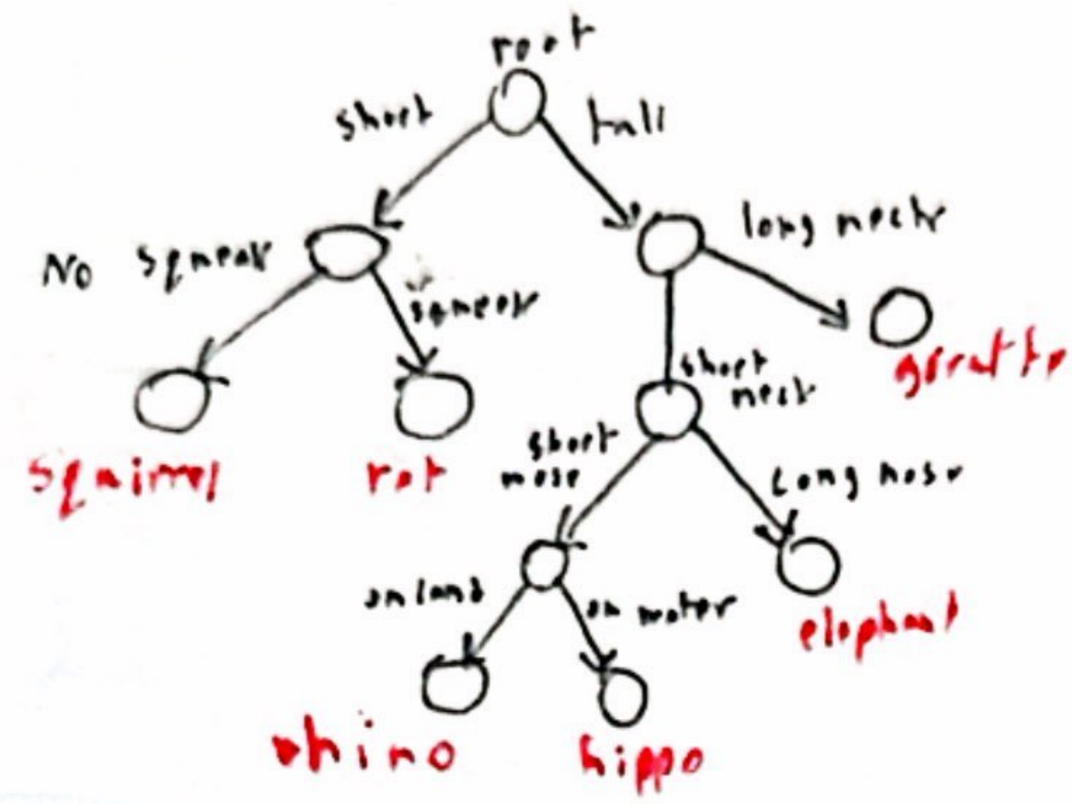
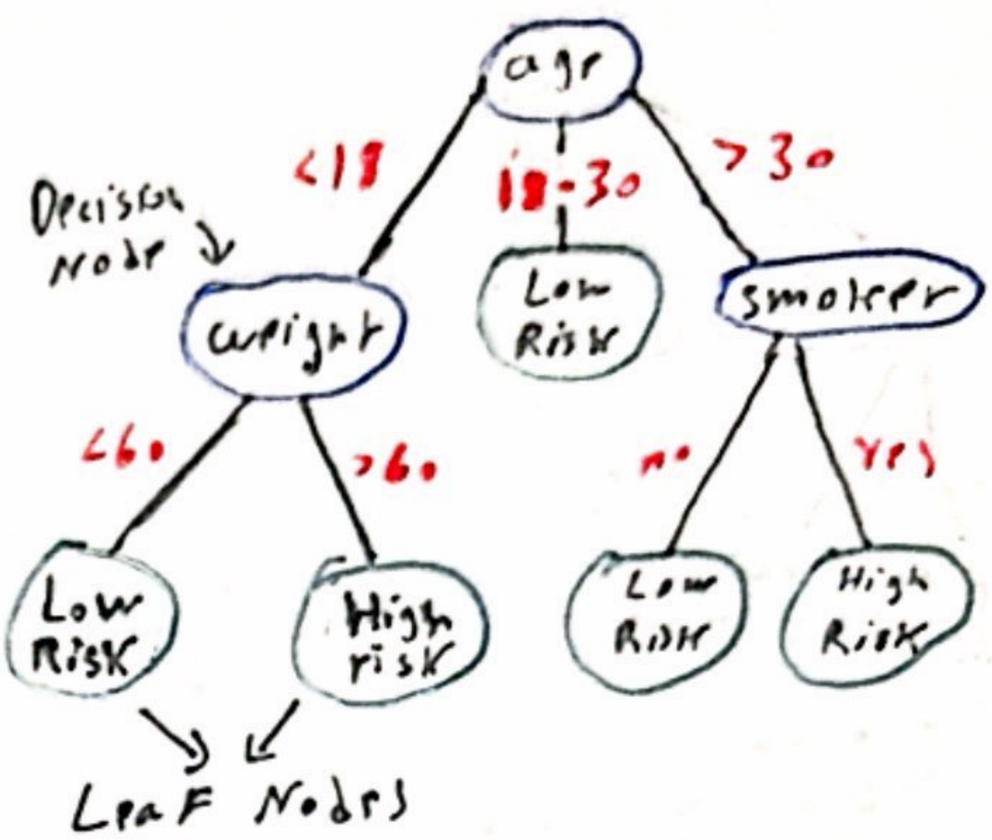
Decision trees (- supervised - classification)

- a tree with a series of Yes No style questions that allow us to partition a data set into several Dimensions

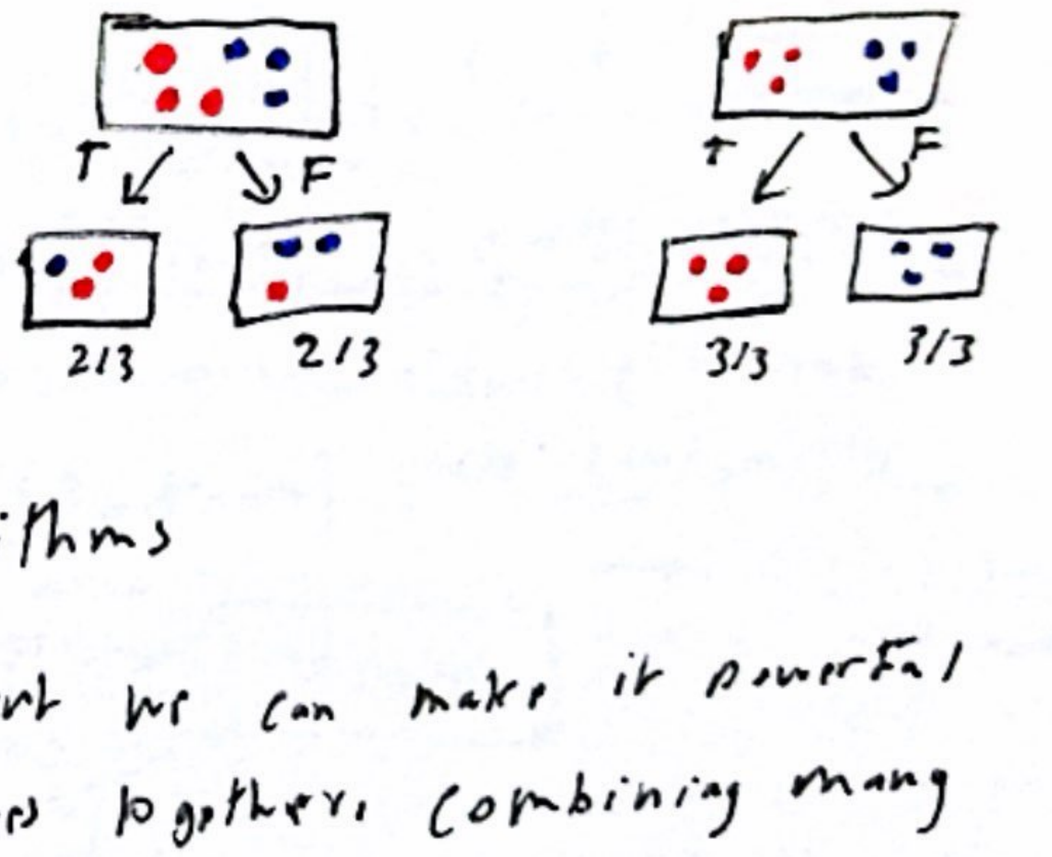
Ex's • Animal classifier



• Heart attack Risk



★ The goal of this algo is to have Leaf nodes that are as pure as possible i.e instead of randomly splitting the data we try to find splits that lead to the resulting groups of leafs to be as pure as possible i.e as few datapoints are misclassified



Ensemble Algorithms

-. a decision tree is simple but we can make it powerful by combining many decision trees together. Combining many simple ML models to a better model is called Ensemble Algorith there are many types of ensembles some discussed in next pages

-Max Depth: is a hyperparameter that limits how many splits (levels) the tree can make from root → Deepest leaf it controls the complexity of tree (how detailed it can be in splitting data). if max Depth too small → tree cannot capture enough patterns — → Underfitting. max Depth too big or unlimited → tree keeps splitting until leaves are pure — → overfit. Right max Depth → good bias variance balance → generalizes well → Good.

once max depth limit is reached node not pure has 24 mixed classes so it chooses the majority class to be the leaf nodes class.