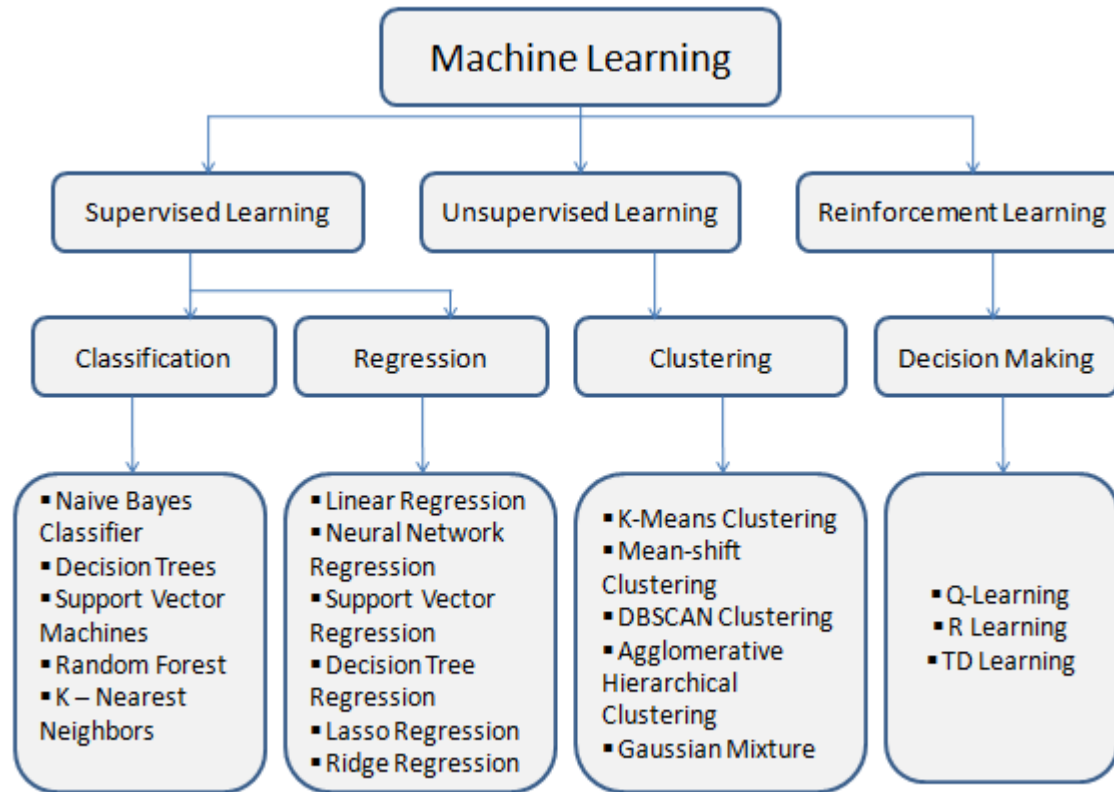


AI Algorithms – 1: Intro to AI Algorithms

- **Sayed Ahmed**
- PhD Studies in Electrical and Computer Eng. (McMaster University) (Partially Complete)
- Master of Engineering in Electrical and Computer Engineering (McMaster University)
- MSc in Data Science and Analytics (Toronto Metropolitan University/Ryerson)
- MSc in Computer Science (U of Manitoba)
- BSc. Engineering in Computer Science and Engineering (BUET)
- Extensive experience in Software Development and Engineering (primarily in Canada)
- Significant experience in Teaching
- Taught in Universities, Colleges, and Training Institutes

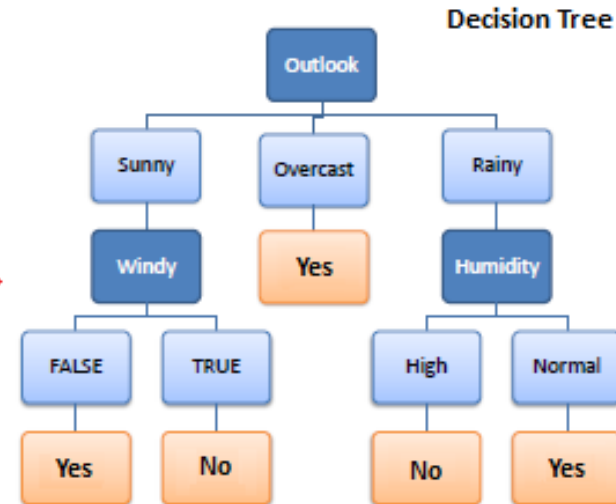


Machine Learning Algorithms

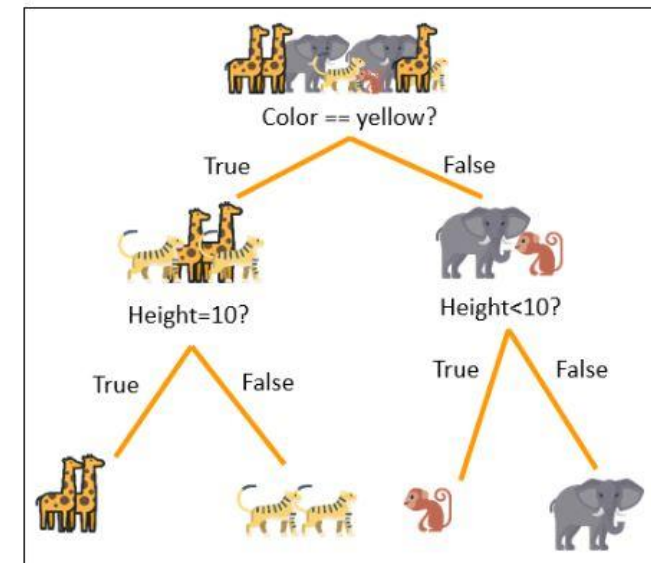


Decision Tree and Classification

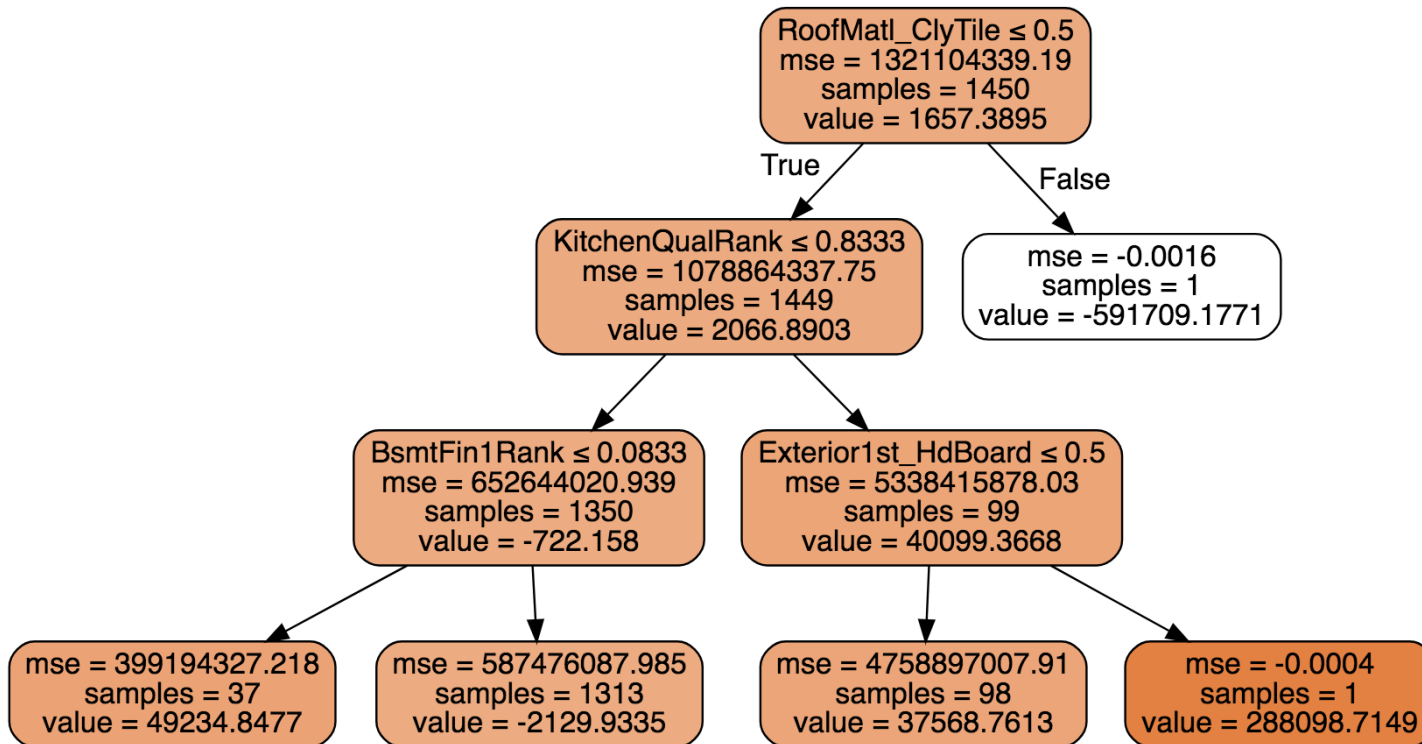
Predictors				Target
Outlook	Temp.	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Google Images, Internet



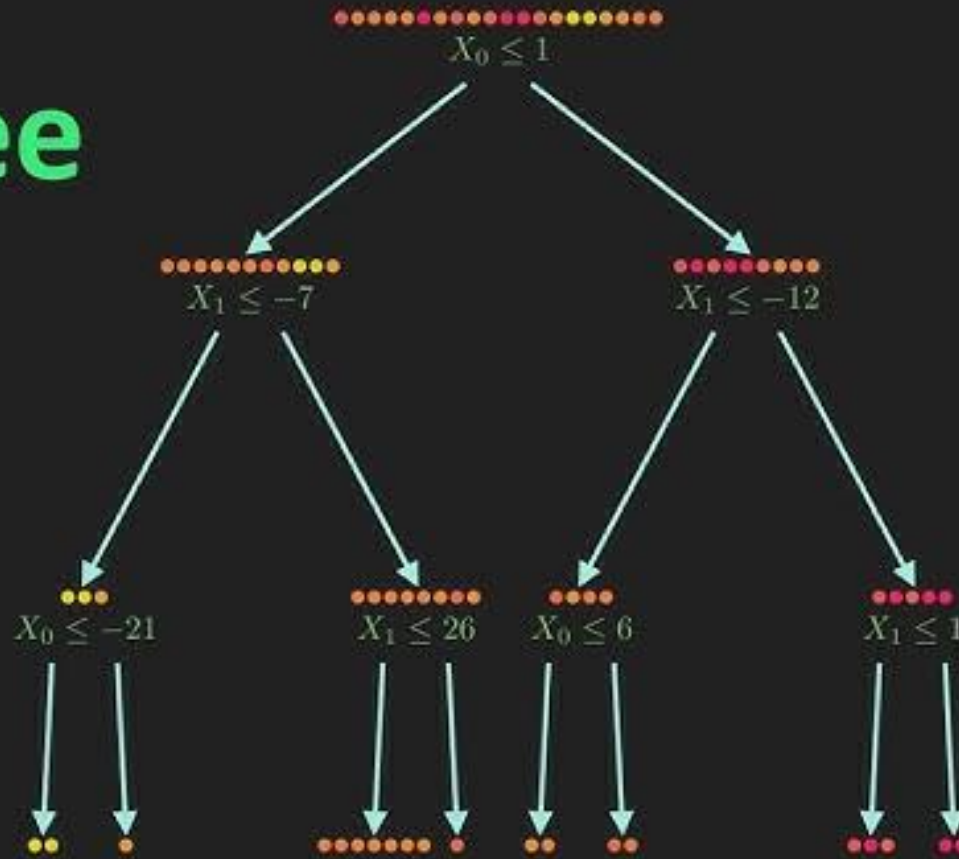
Decision Tree and Regression



Regression and Decision Trees

Decision Tree Regression

From Scratch

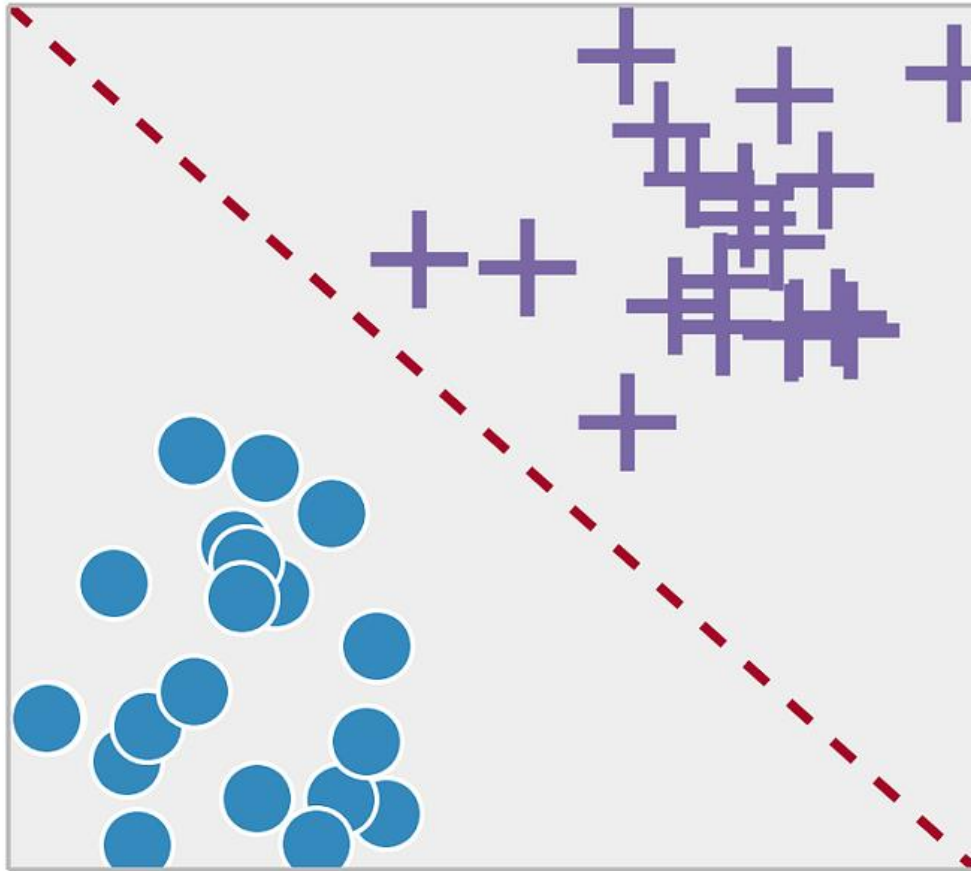


Google Images, Internet

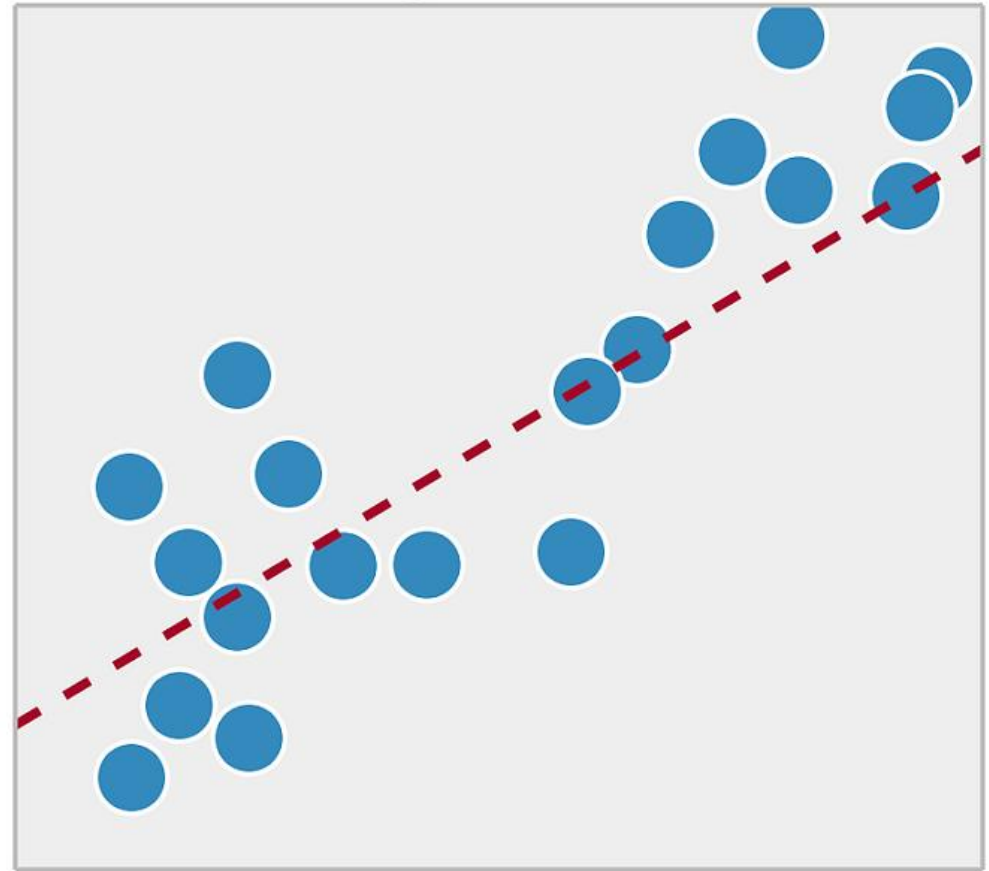
Google Images, Internet

Supervised Learning

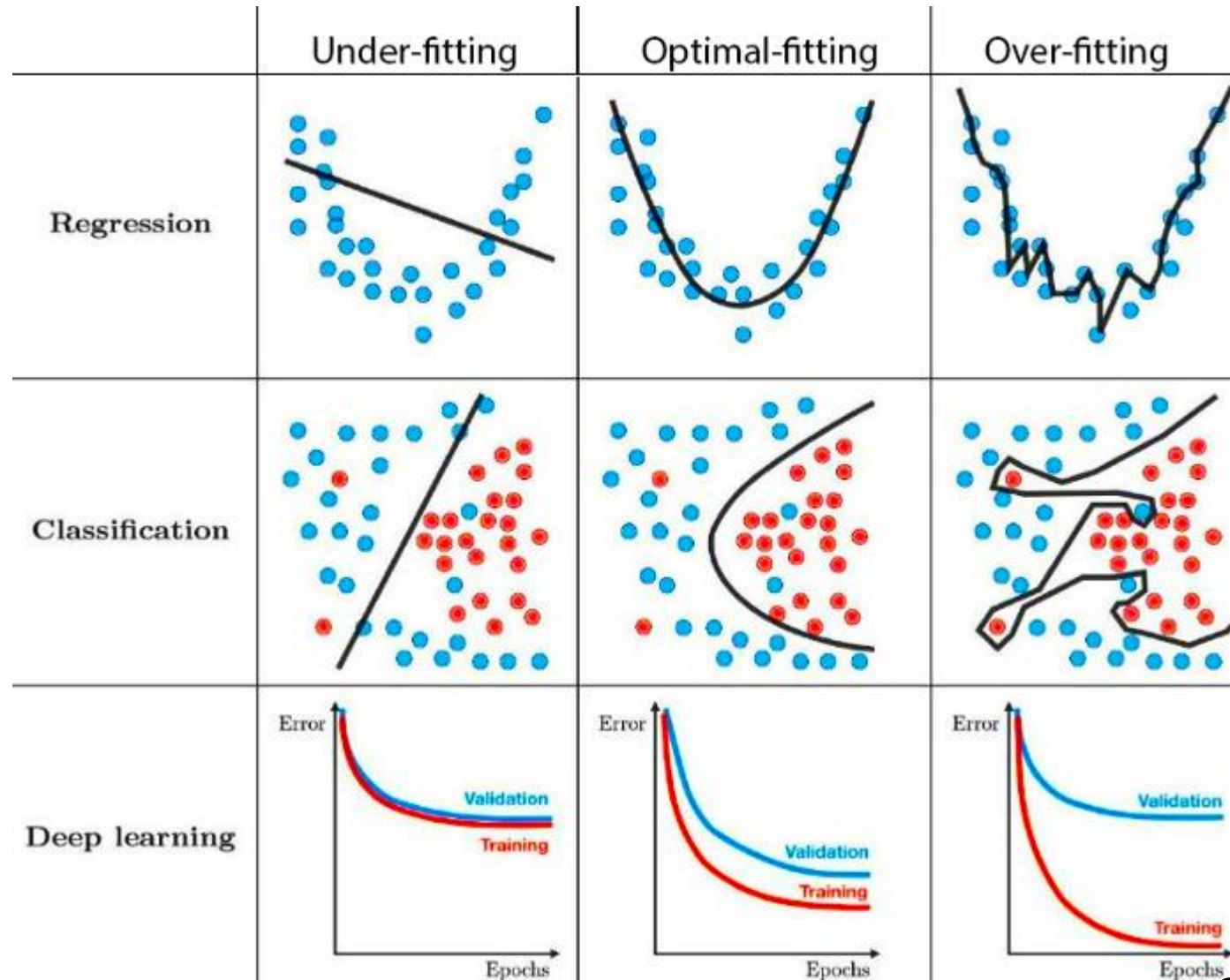
Classification



Regression



Underfit, Overfit, Optimal Fit

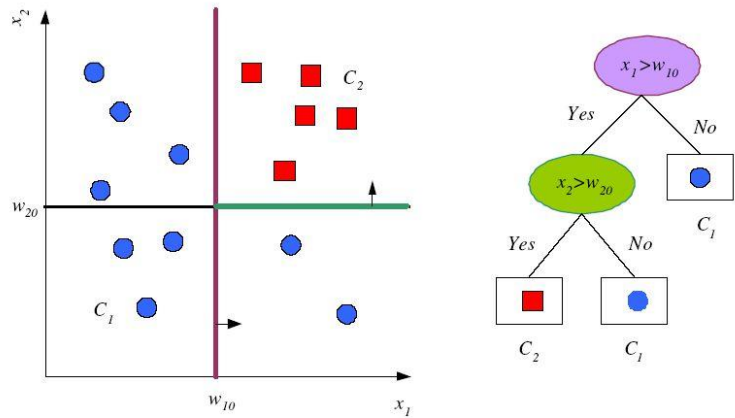


Google Images, Internet

Decision Tree Approaches



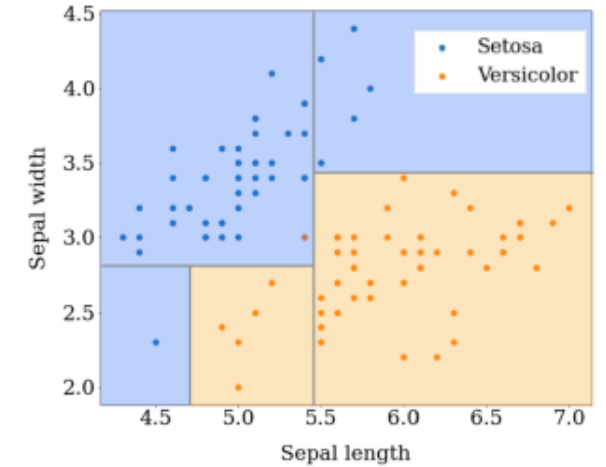
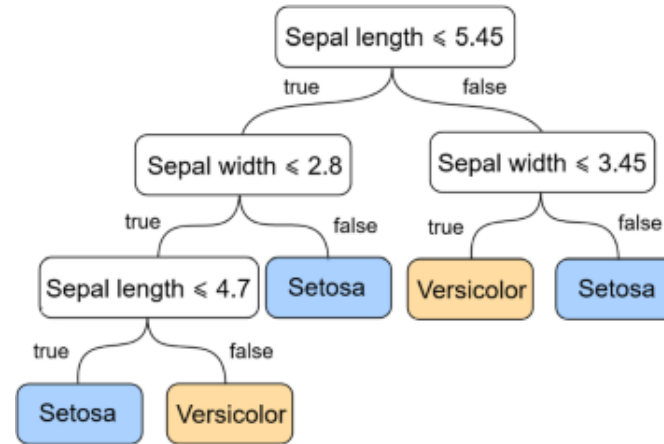
Tree Uses Nodes, and Leaves



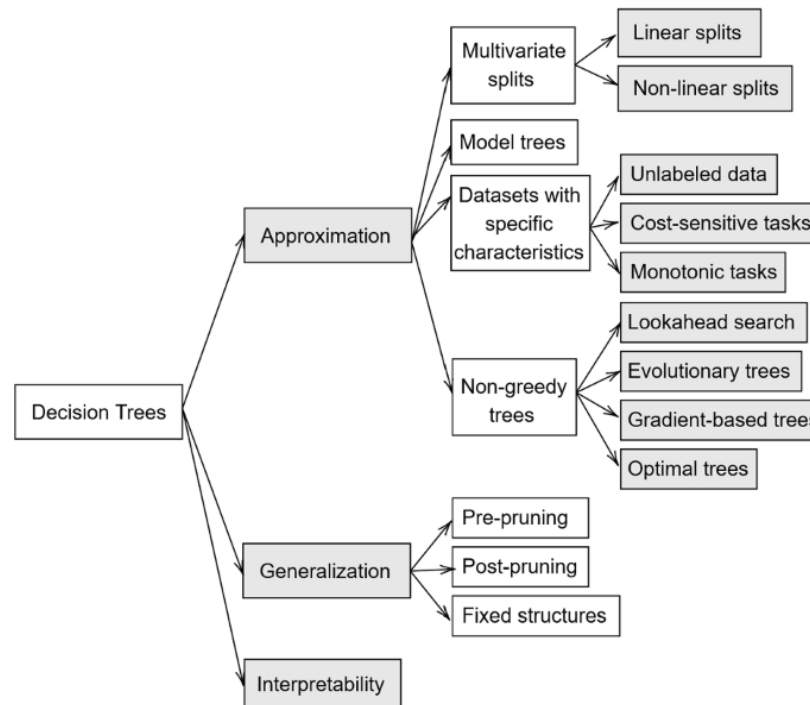
Lecture Notes for E Alpaydm 2004 Introduction to Machine Learning © The MIT Press (V1.1)

2

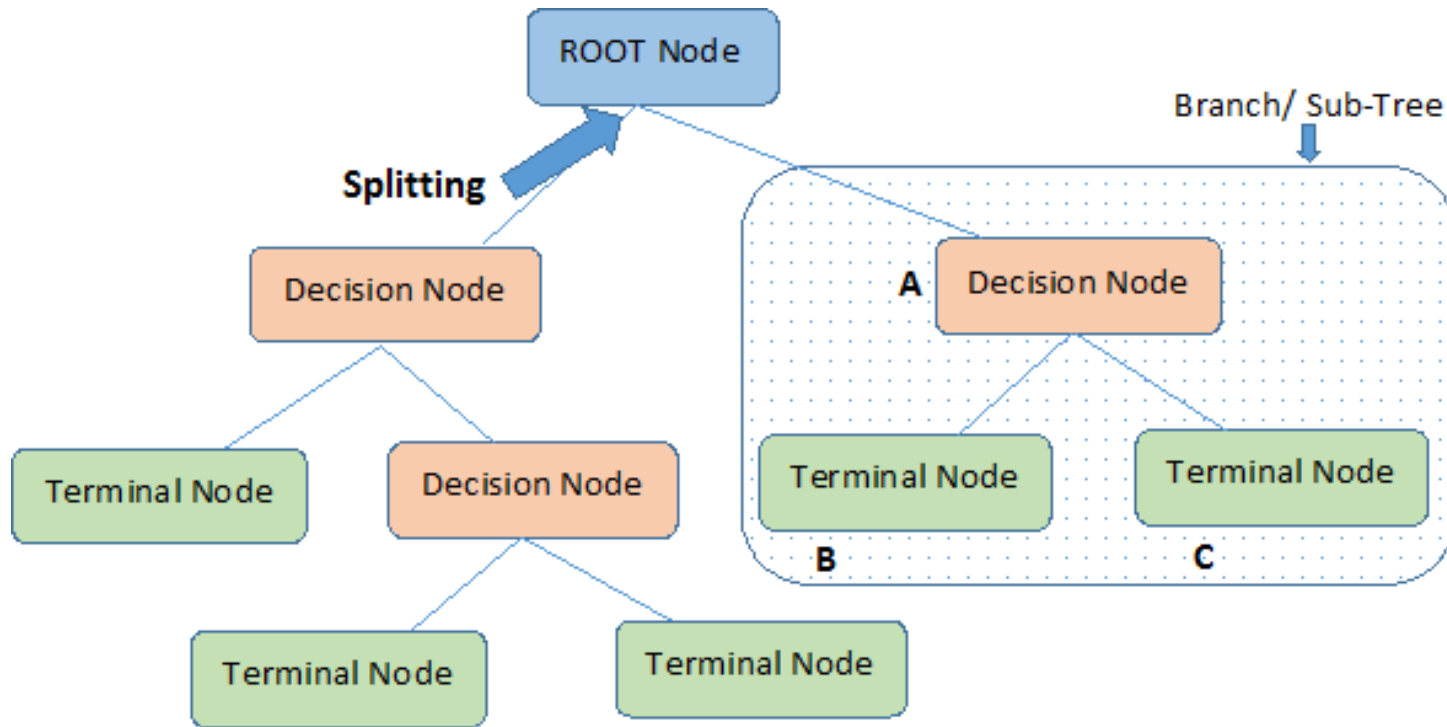
Google Images, Internet



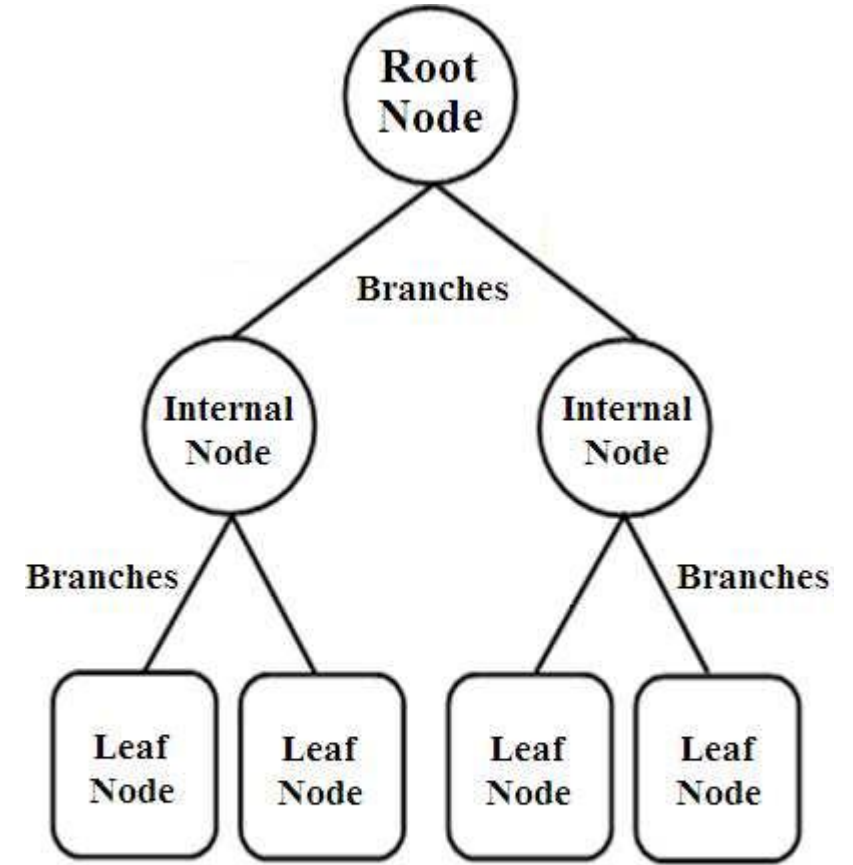
(b) Partitioning visualization



Decision Tree, Node Types



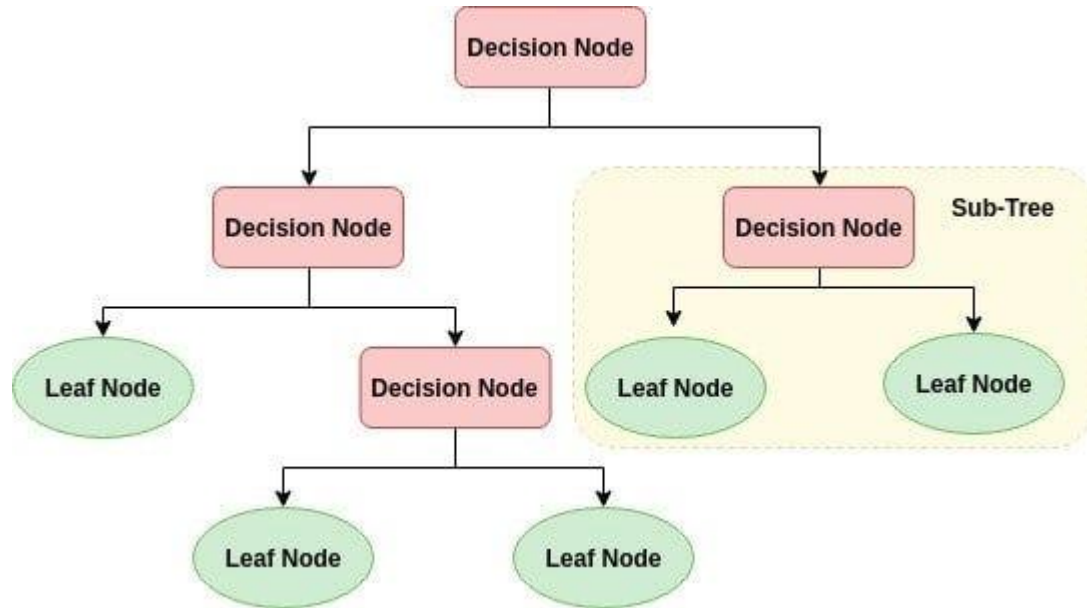
Note:- A is parent node of B and C.



(a)

Google Images, Internet

Node Types



Google Images, Internet

Pros and Cons: Decision Trees

Decision Tree Pros and Cons

Pros

- No data assumptions
- Non-linear
- Discontinuous

Cons

- No extrapolation
- Need substantial amount of data
- Not descriptive

DECISION TREES

ADVANTAGES AND DISADVANTAGES

<p>#1. VERY EASY TO EXPLAIN! EVEN EASIER THAN LINEAR REGRESSION.</p> <p>#2. MORE CLOSELY MIRROR HUMAN DECISION-MAKING THAN OTHER REGRESSION AND CLASSIFICATION APPROACHES.</p> <p>#3. CAN EASILY HANDLE QUALITATIVE PREDICTORS WITHOUT THE NEED TO CREATE DUMMY VARIABLES.</p>	<p>#1. DO NOT HAVE THE SAME LEVEL OF PREDICTIVE ACCURACY AS SOME OTHER REGRESSION AND CLASSIFICATION APPROACHES.</p> <p>#2. CAN BE VERY NON-ROBUST... A SMALL CHANGE IN THE DATA CAN CAUSE A LARGE CHANGE IN THE LARGE ESTIMATED TREE.</p>
--	--

lg

Pros and cons of decision tree analysis

Pros



Transparent

Offers a clear method for you to make decisions.



Efficient

Requires little time and few resources for you to create.



Flexible

Allows you to add decisions to the tree if needed.

Cons



Complex

Can become complex if you add too many decisions.



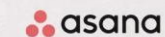
Unstable

Can become unstable if you change your data.



Risky

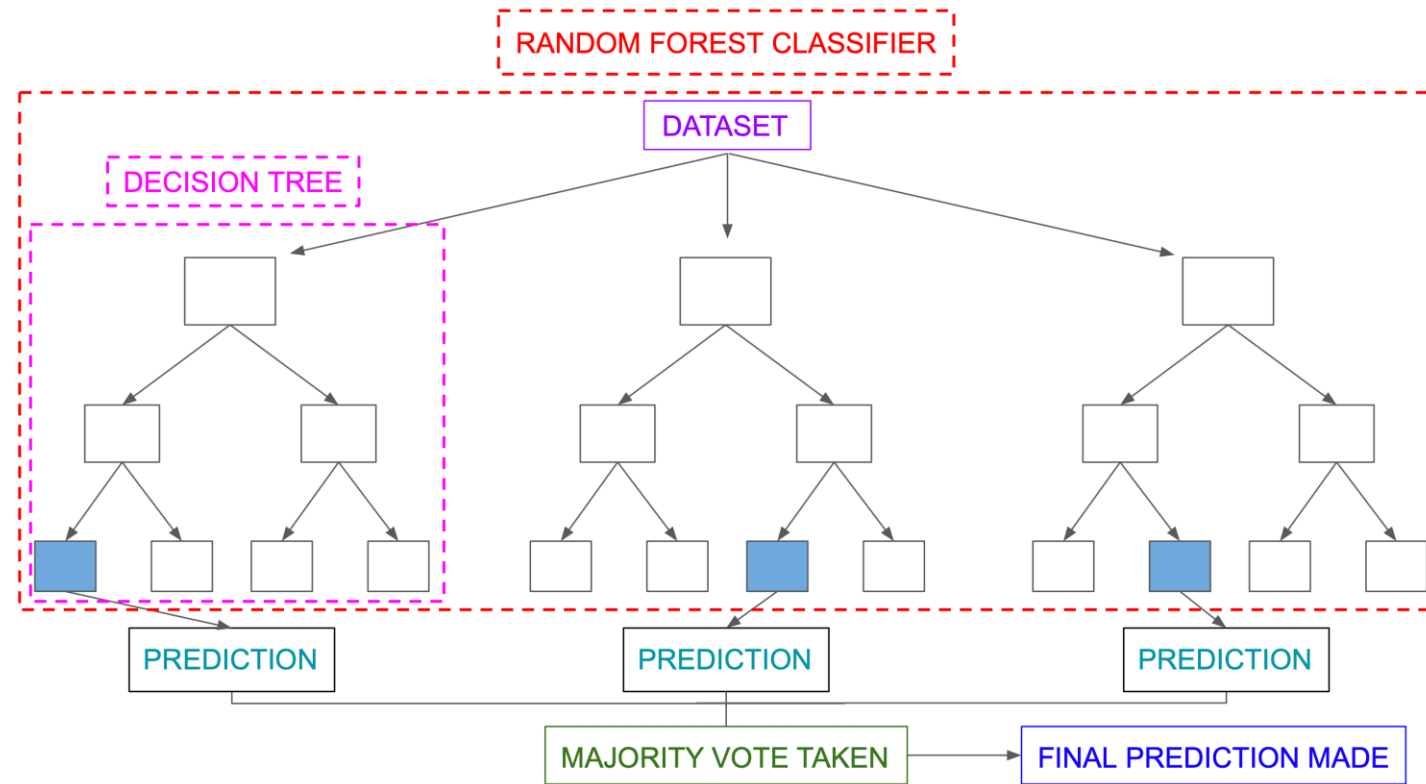
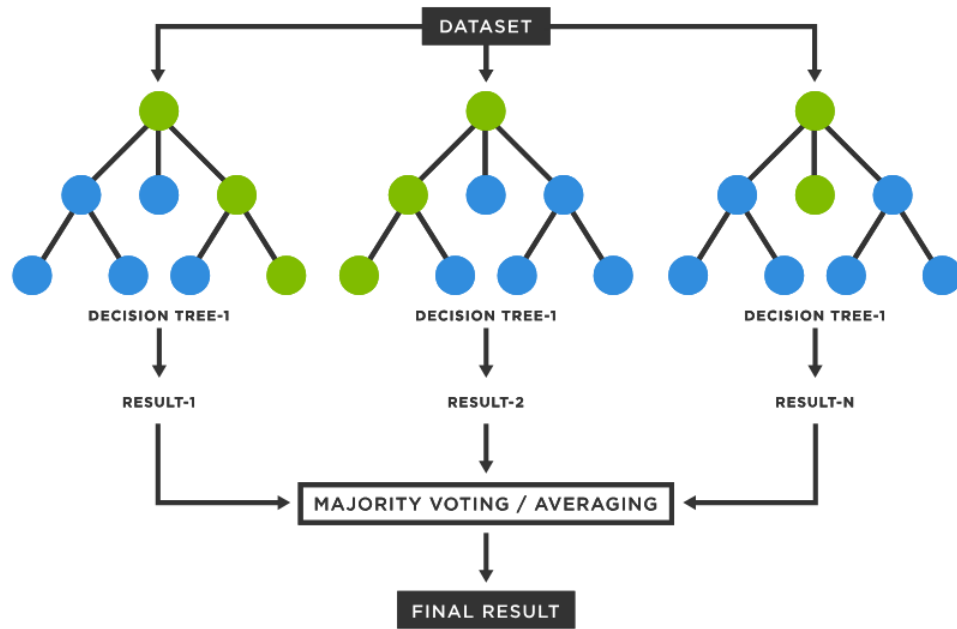
Can become risky if you don't analyze possible outcomes.



Decision Trees vs Random Forest

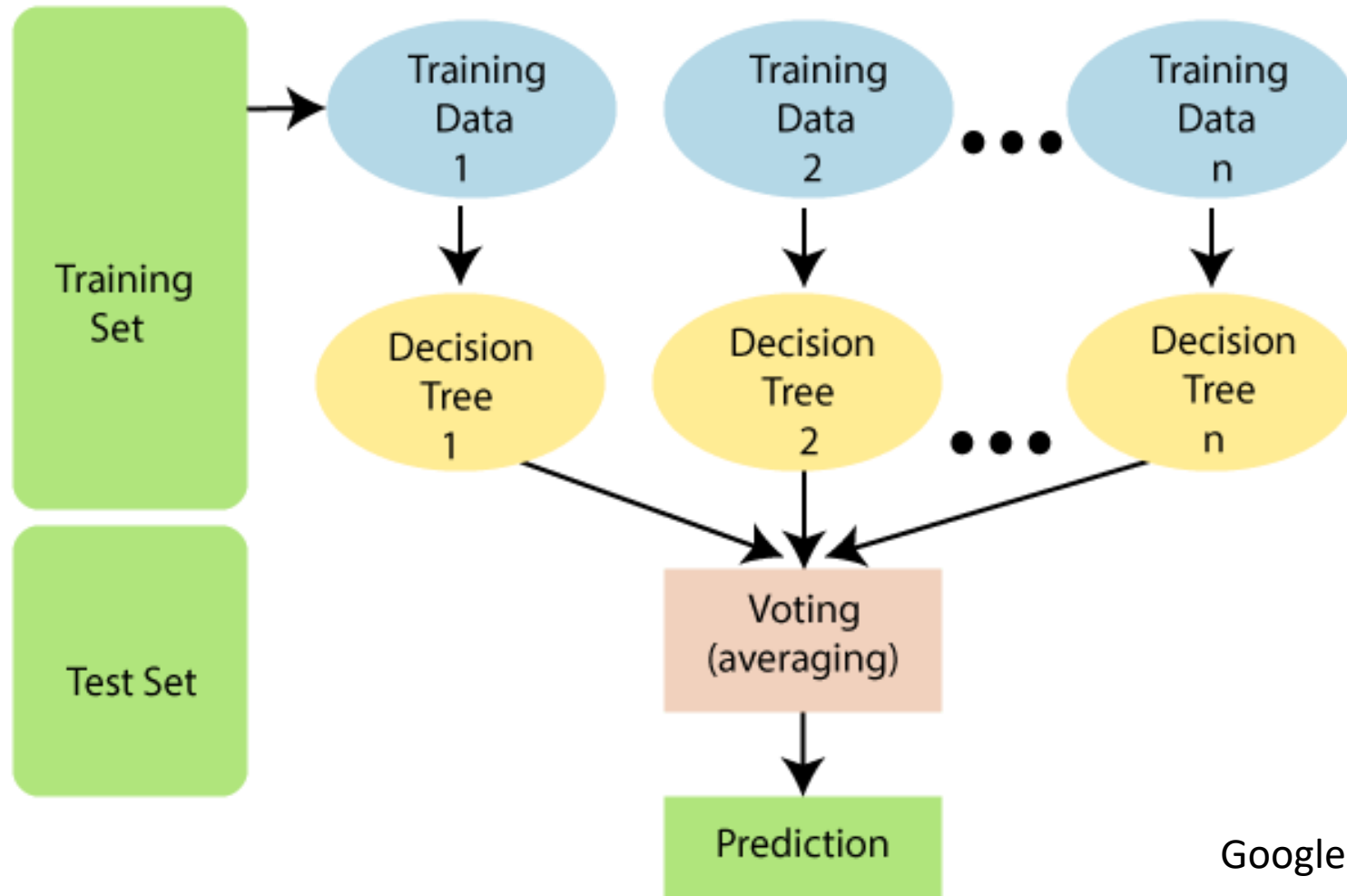
	Decision Tree	Random Forest
Interpretability	Easy to interpret	Hard to interpret
Accuracy	Accuracy can vary	Highly accurate
Overfitting	Likely to overfit data	Unlikely to overfit data
Outliers	Can be highly affected by outliers	Robust against outliers
Computation	Quick to build	Slow to build (computationally intensive)

Random Forest



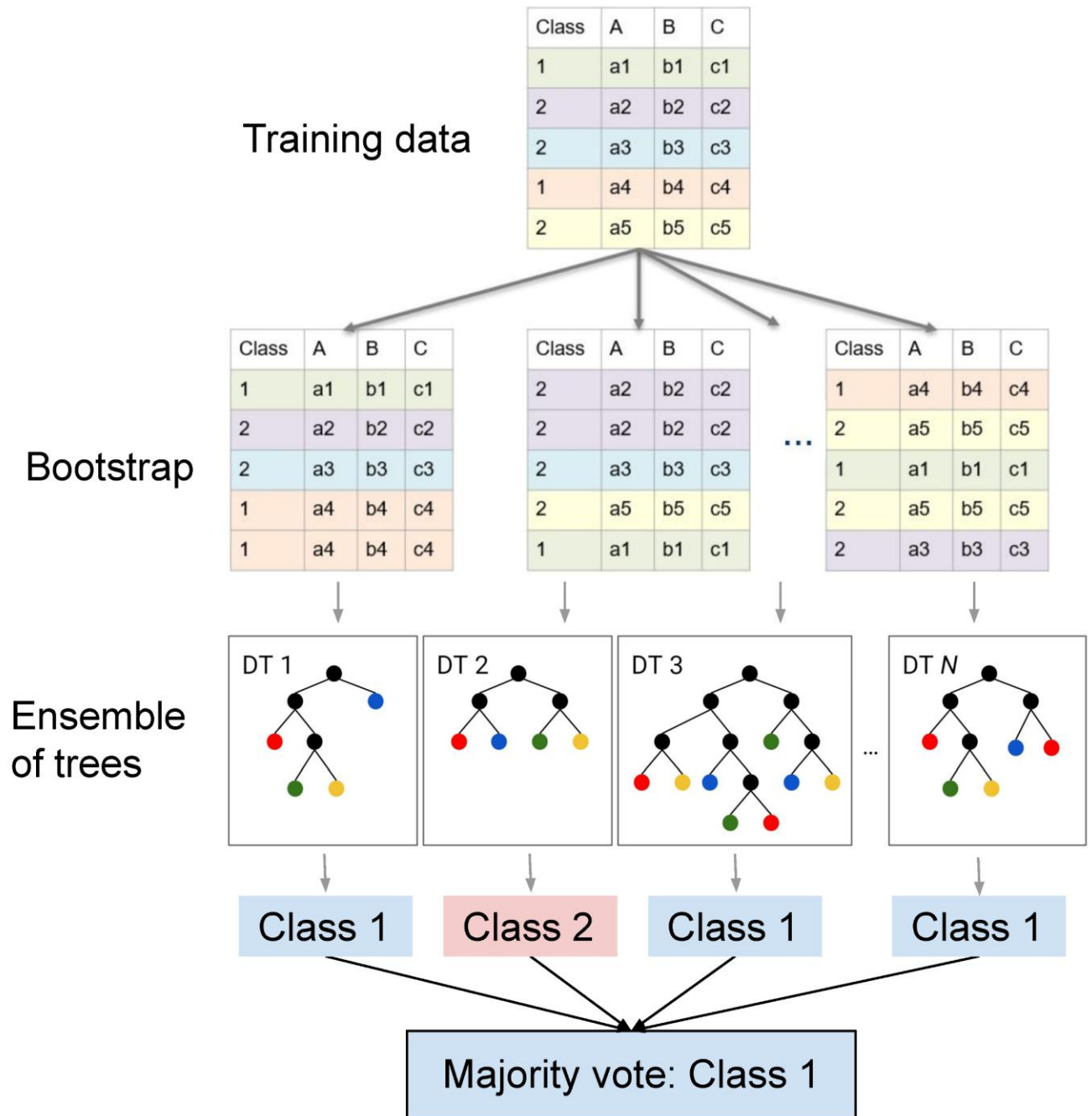
Google Images, Internet

Random Forest



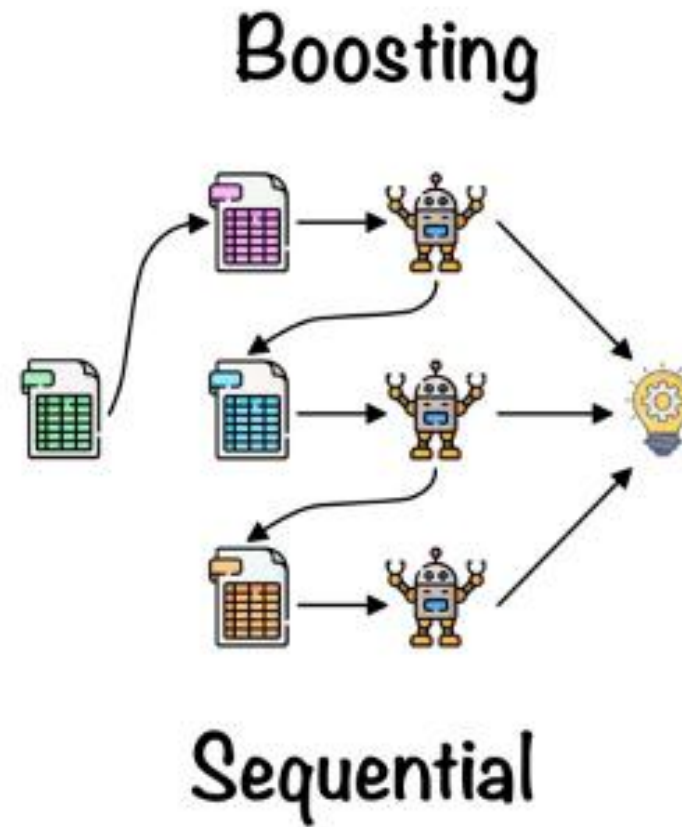
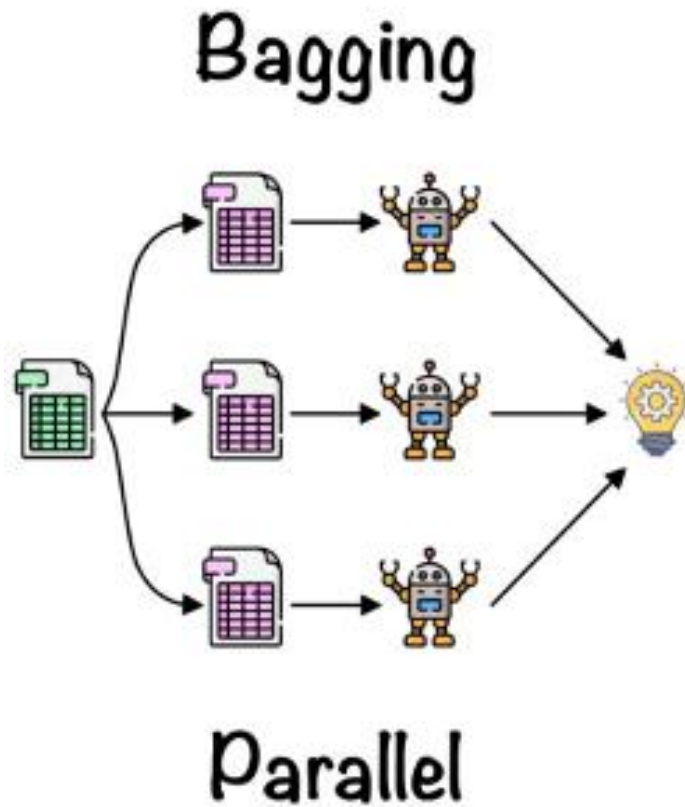
Google Images, Internet

Random Forest



Google Images, Internet

Random Forest Algorithms

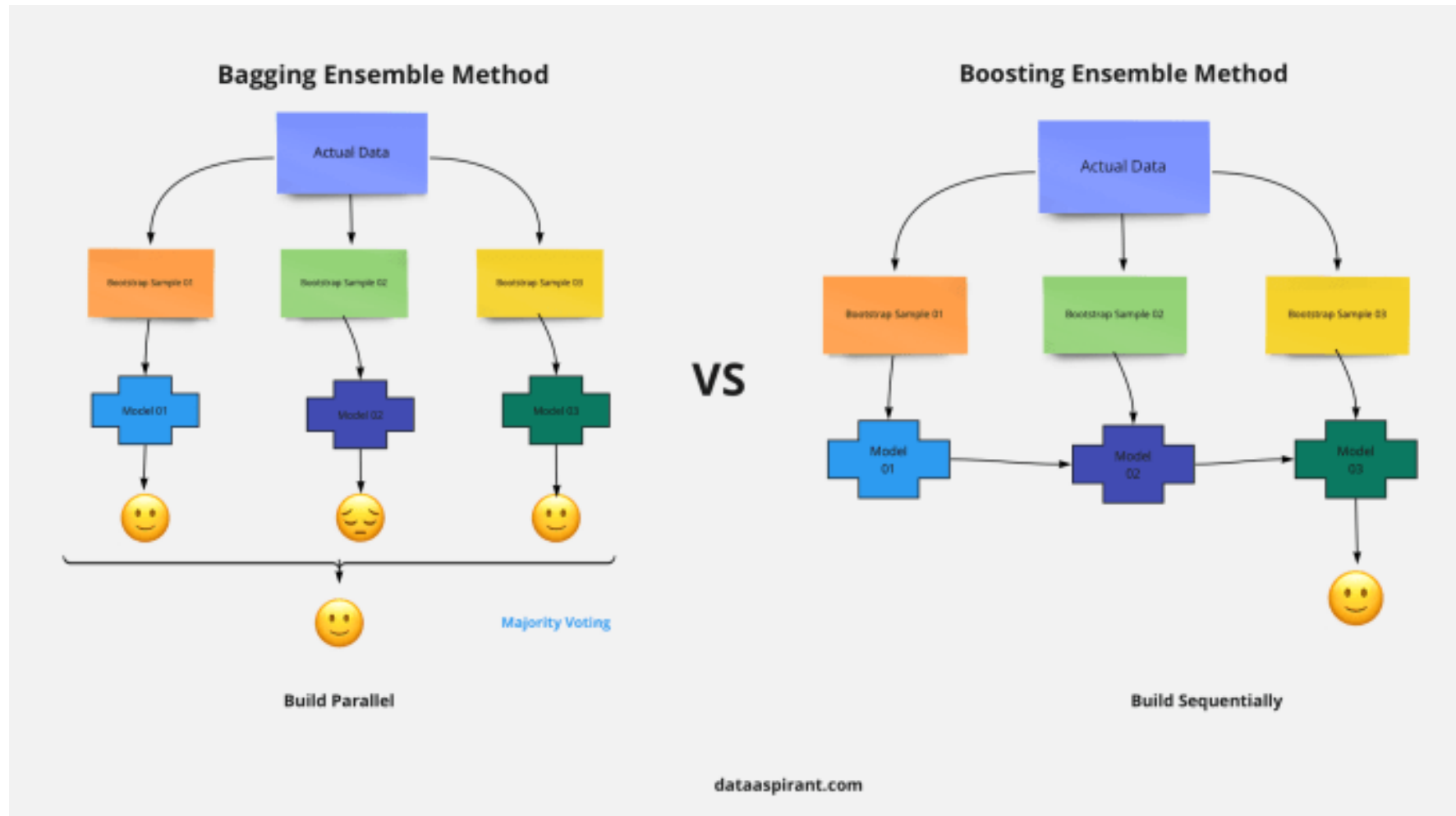


Bagging, Boosting, Stacking

	Bagging	Boosting	Stacking
Purpose	Reduce Variance	Reduce Bias	Improve Accuracy
Base Learner Types	Homogeneous	Homogeneous	Heterogeneous
Base Learner Training	Parallel	Sequential	Meta Model
Aggregation	Max Voting, Averaging	Weighted Averaging	Weighted Averaging

Google Images, Internet

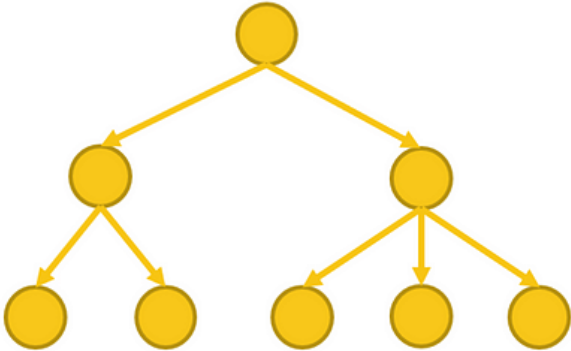
Ensemble Methods



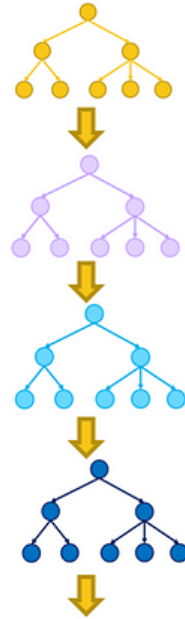
Google Images, Internet

Gradient Boosted Trees

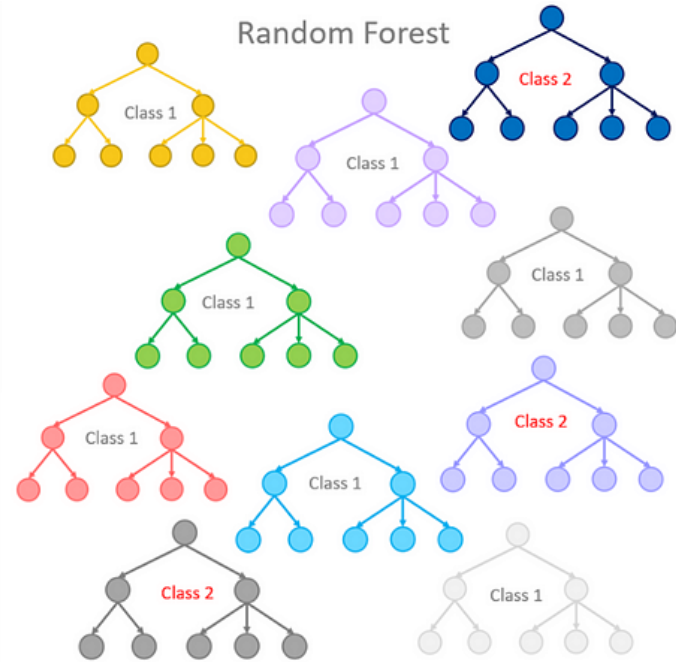
Single Decision Tree



Gradient Boosted Trees



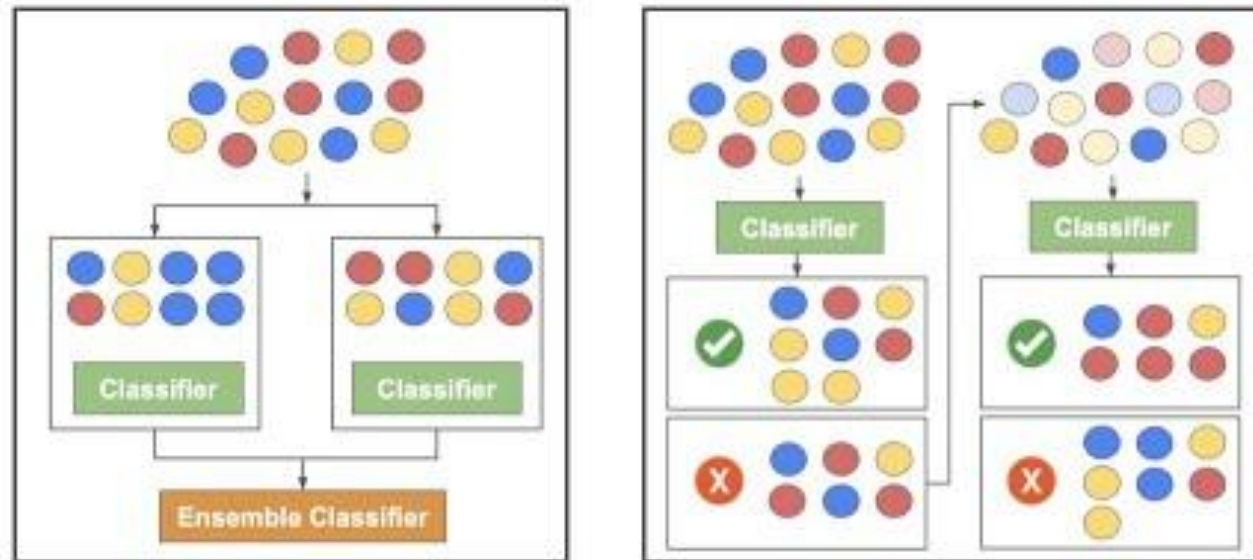
Random Forest



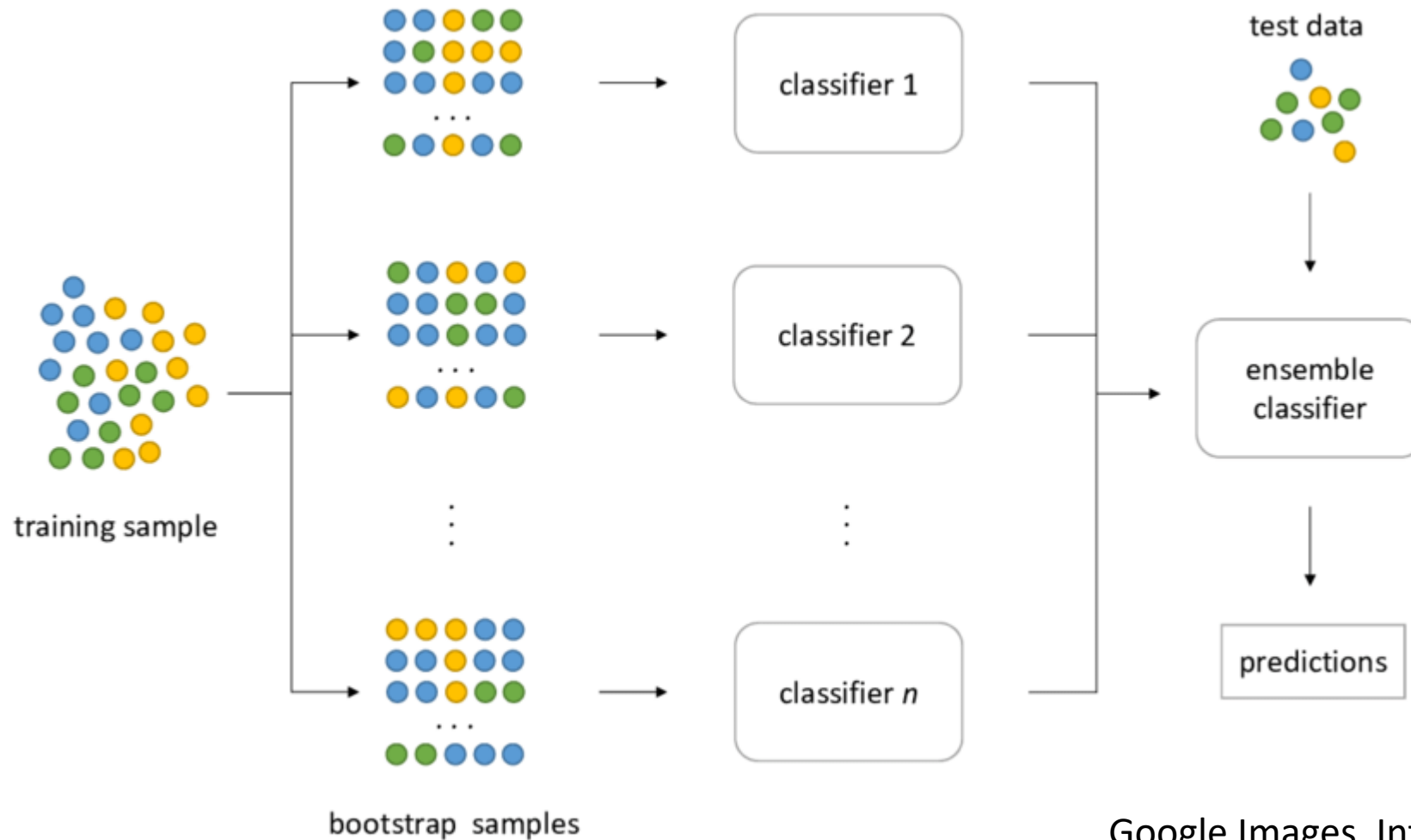
Google Images, Internet

Bagging vs Boosting

Bagging vs Boosting



Bootstrapping



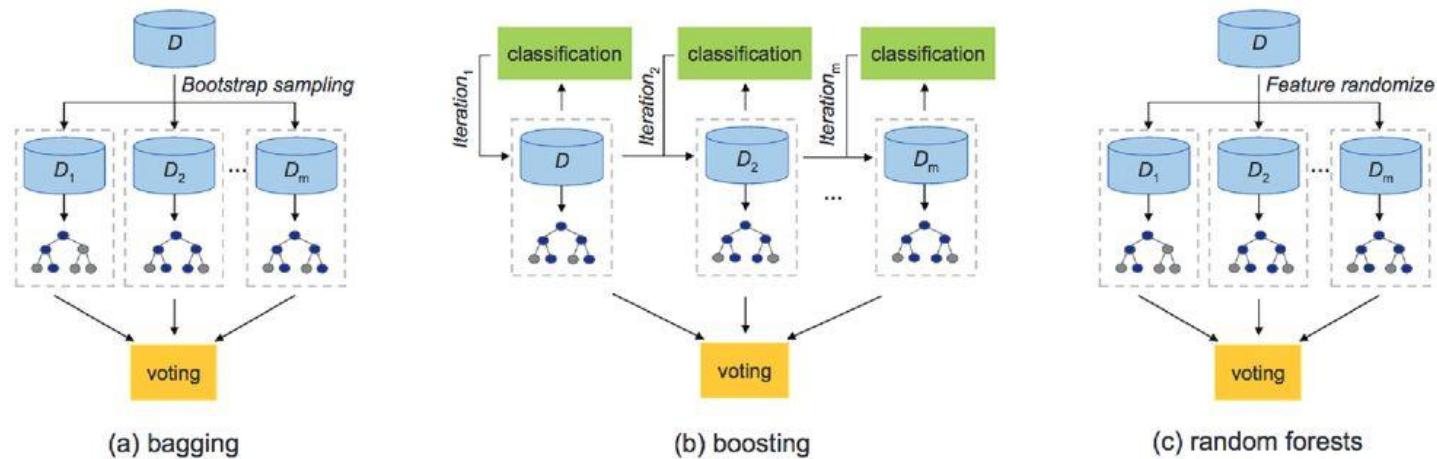
Google Images, Internet

Ensemble Learning

Ensemble learning

Aggregating a group of classifiers (“base classifiers”) as an ensemble committee and making the prediction by consensus.

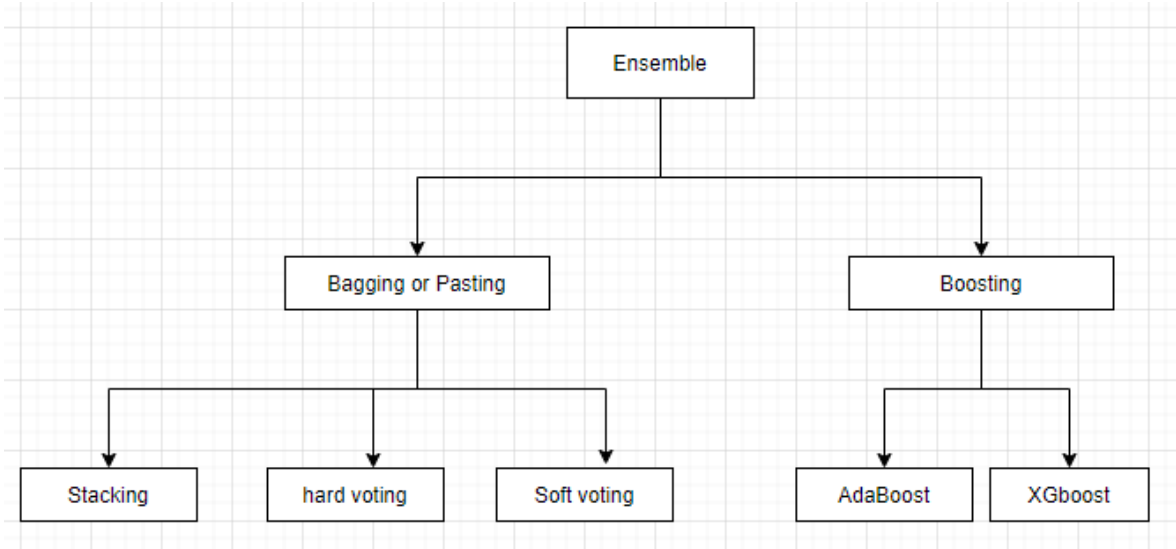
Weak learner ensembles (each base learner has high EPE, but is easy to train):



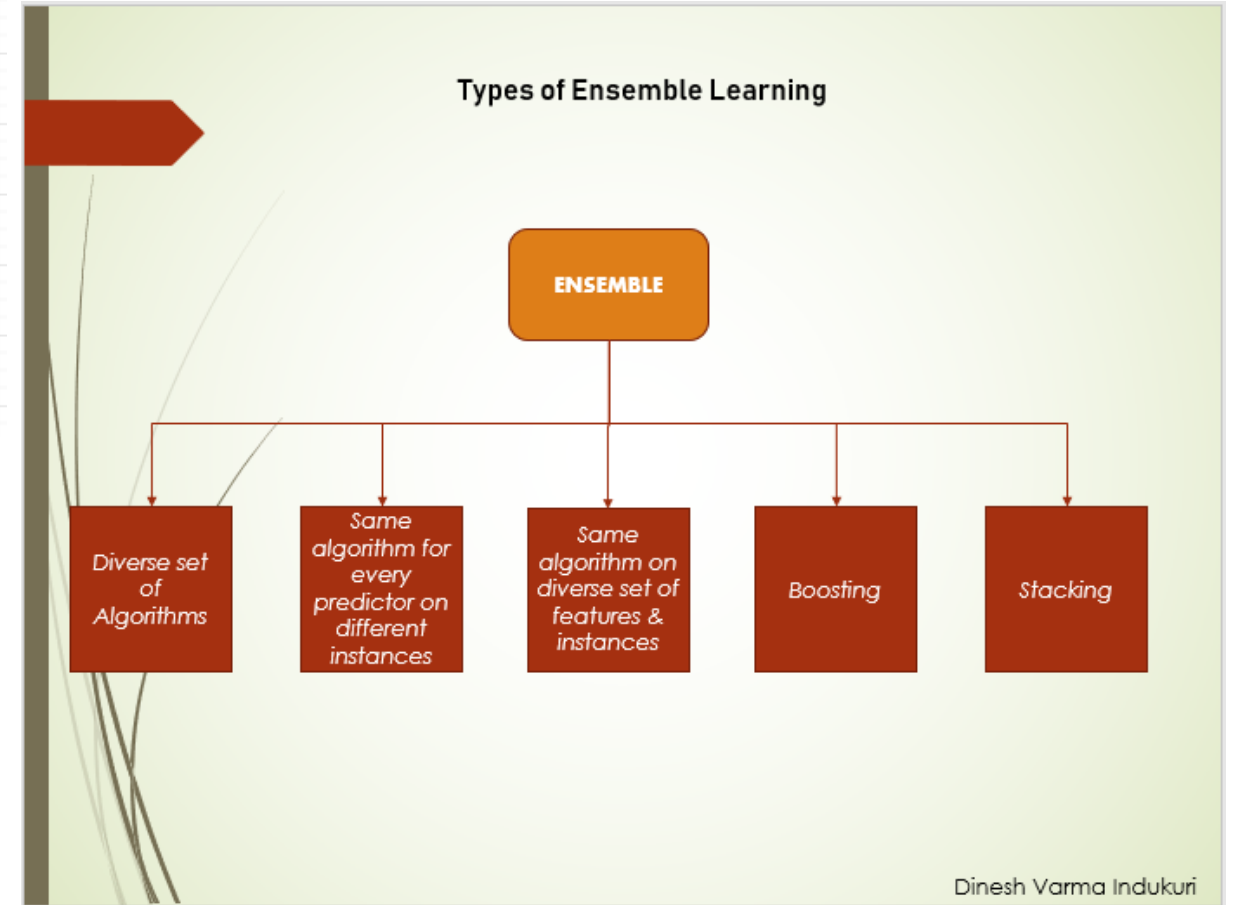
Google Images, Internet

Fig. 1: Schematic illustration of the three popular ensemble methods.

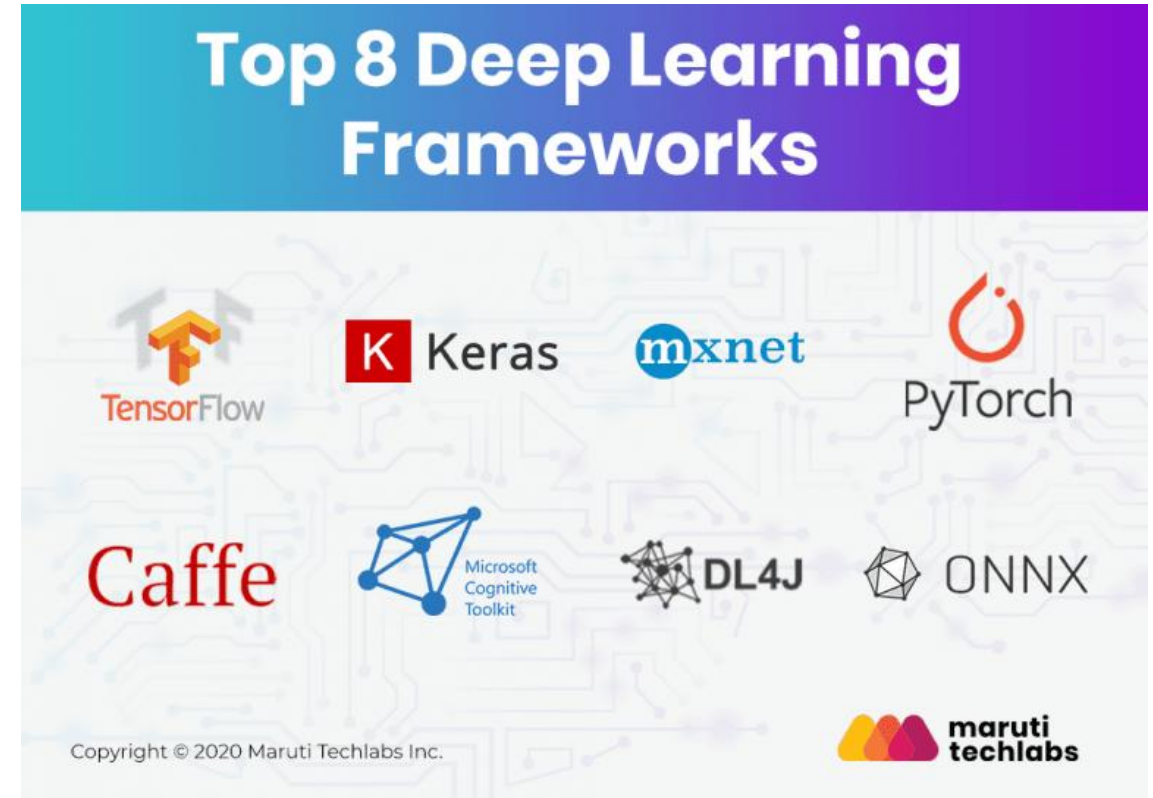
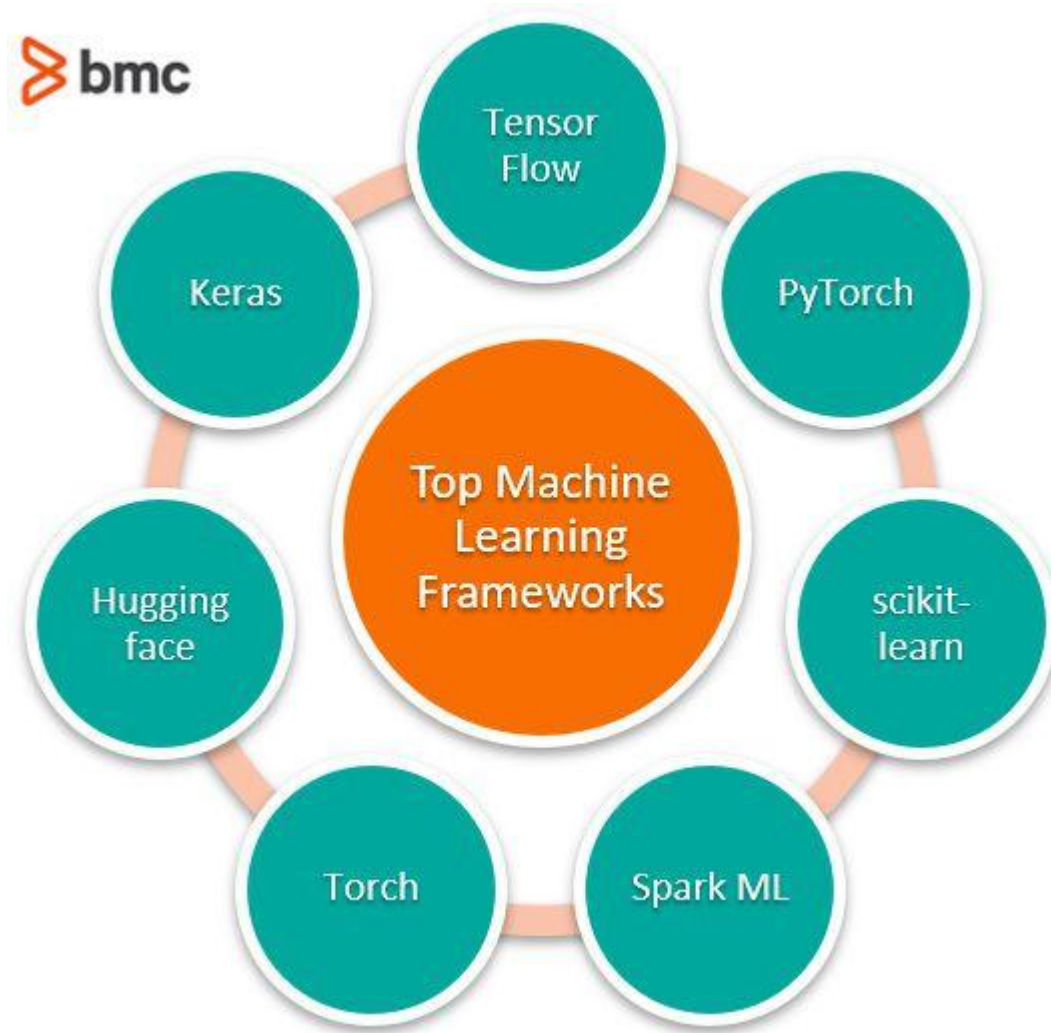
Types of Ensemble Learning



Google Images, Internet



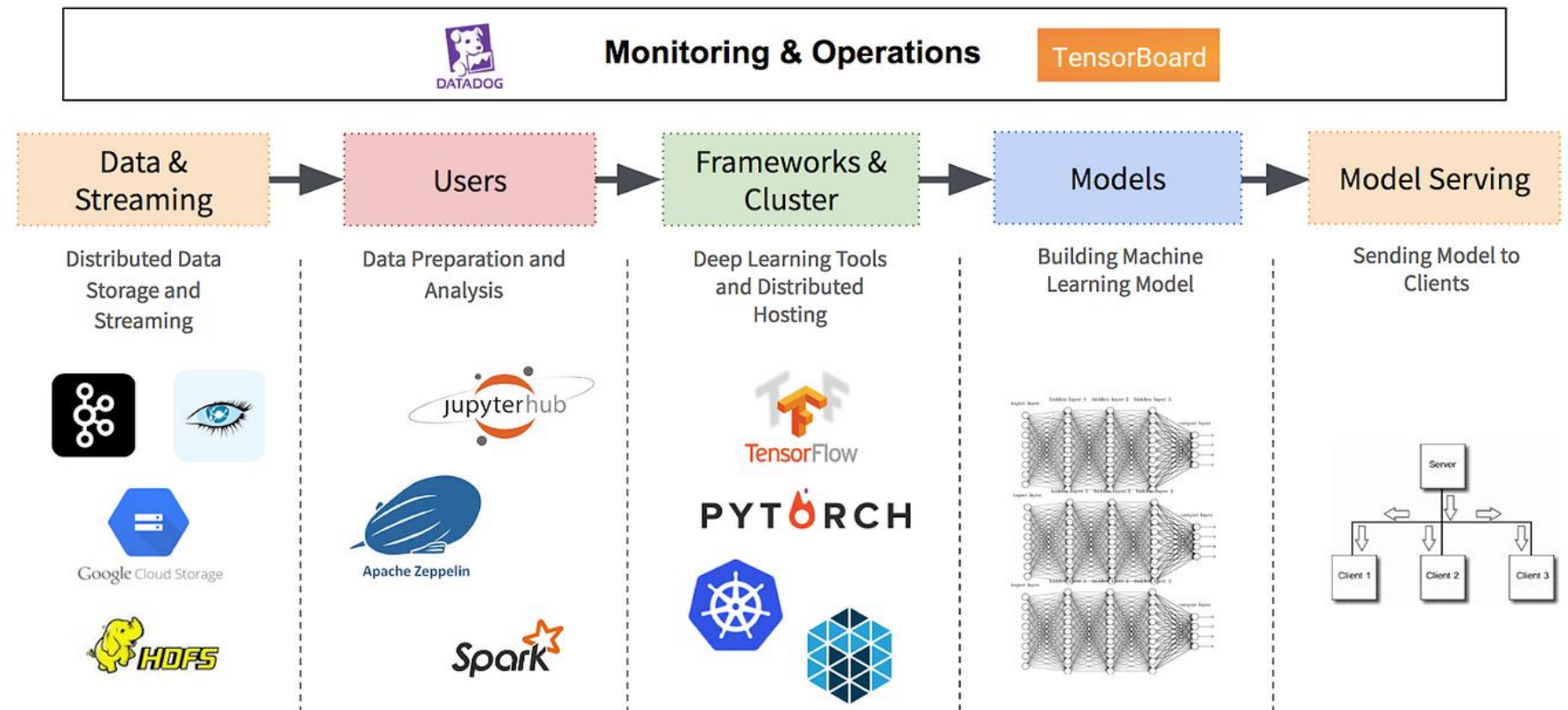
Machine Learning Frameworks



Google Images, Internet

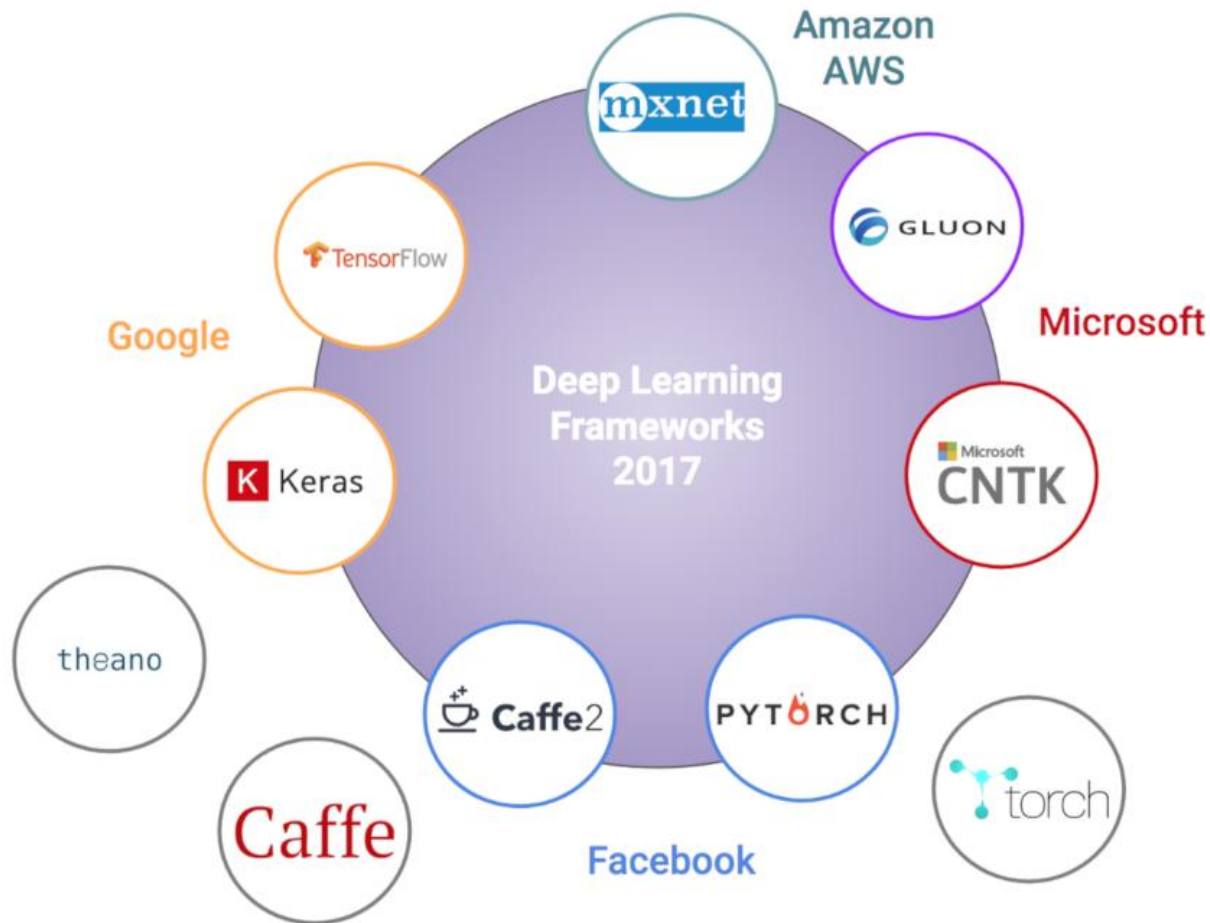
Machine Learning Frameworks

Deep Learning Pipeline



Google Images, Internet

Machine Learning Frameworks



Google Images, Internet