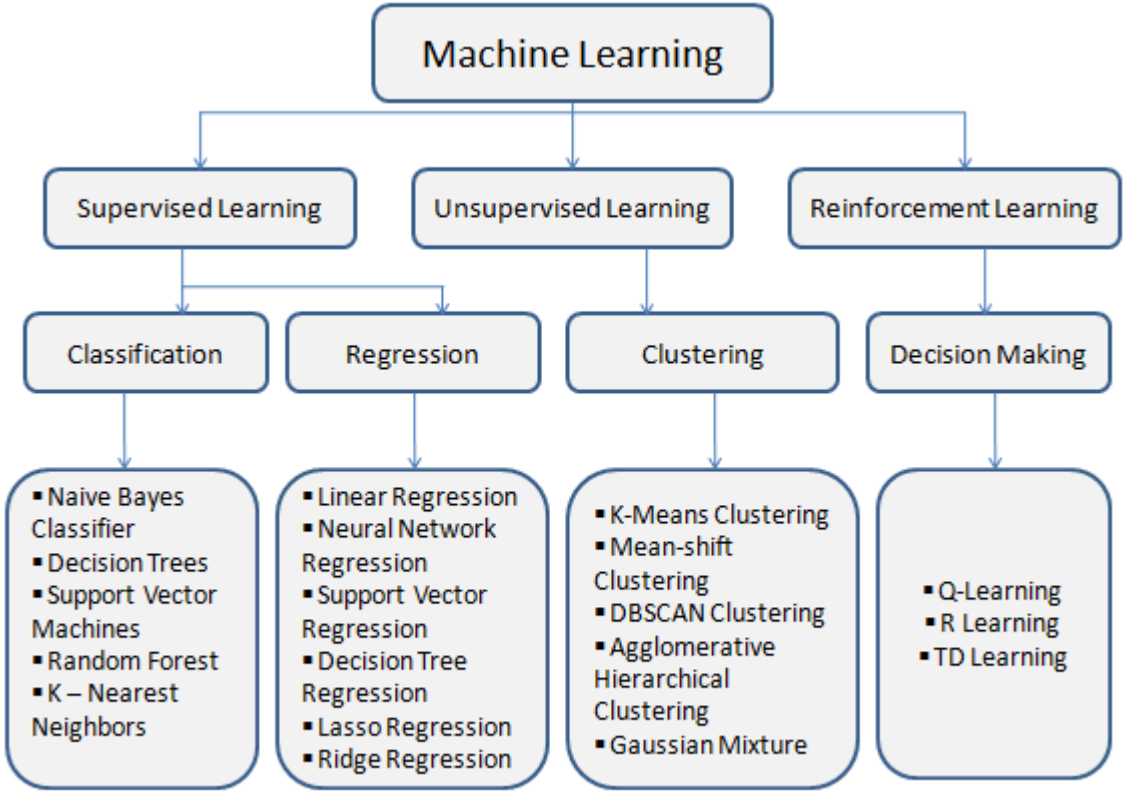
AI Algorithms – 1: Intro to AI Algorithms

- **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

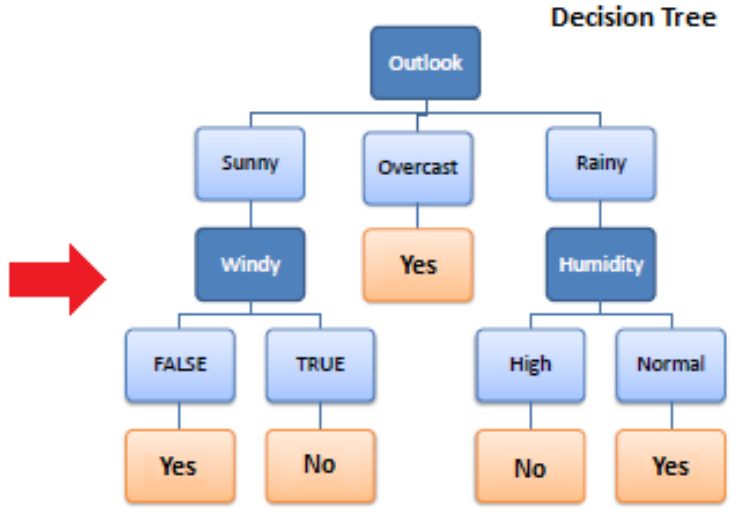
Machine Learning Algorithms



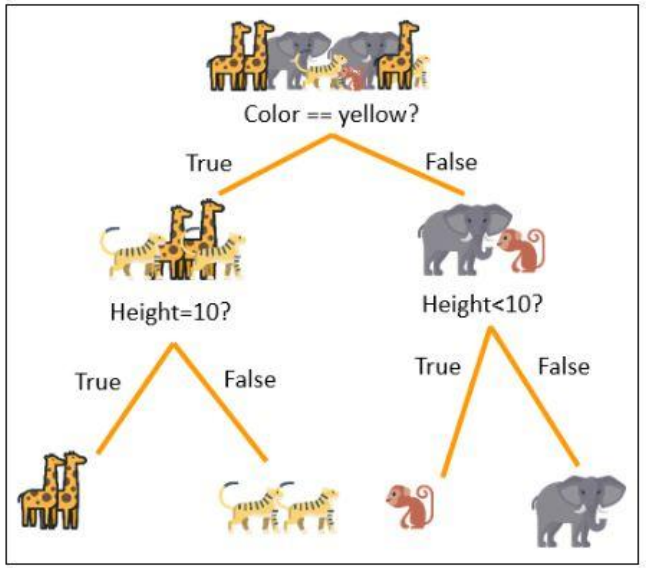
Google Images, Internet

Decision Tree and Classification

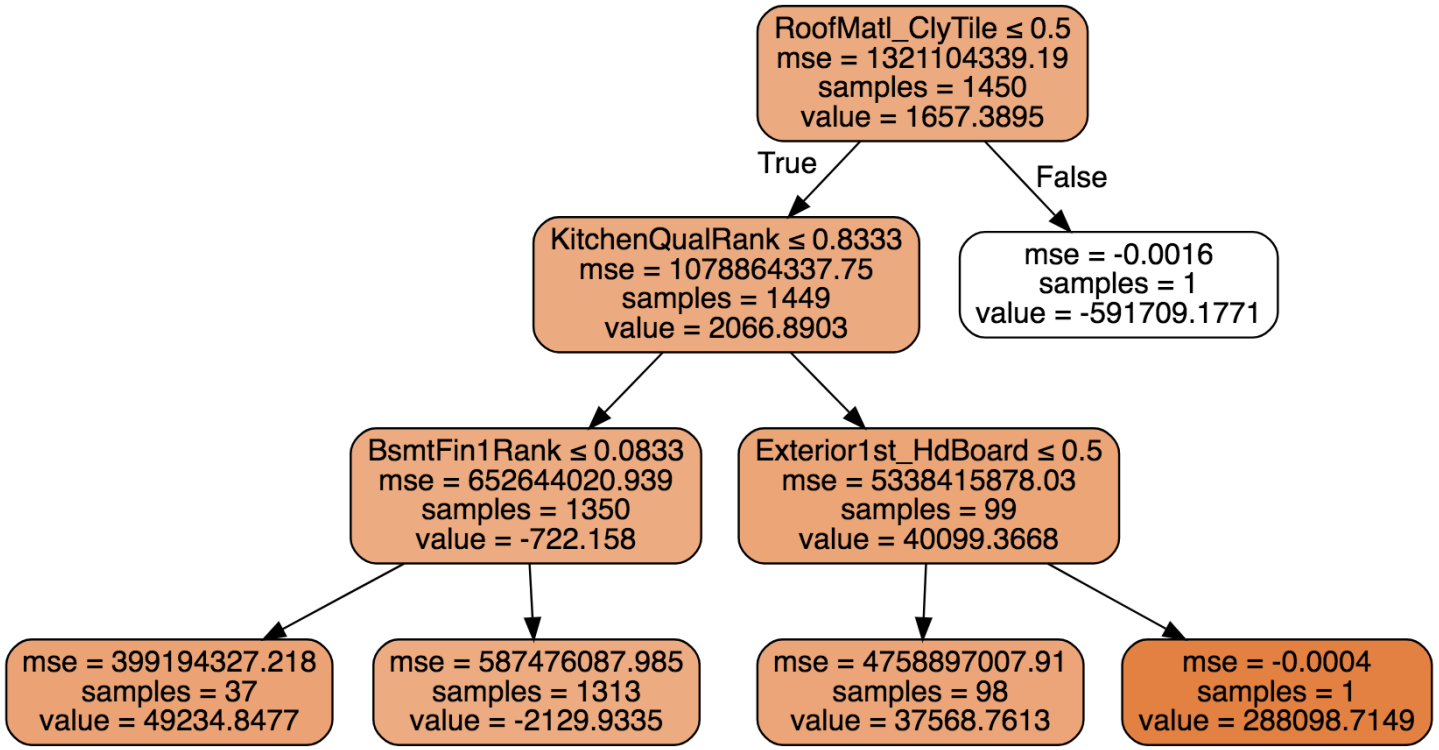
Predictors				Target
Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny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

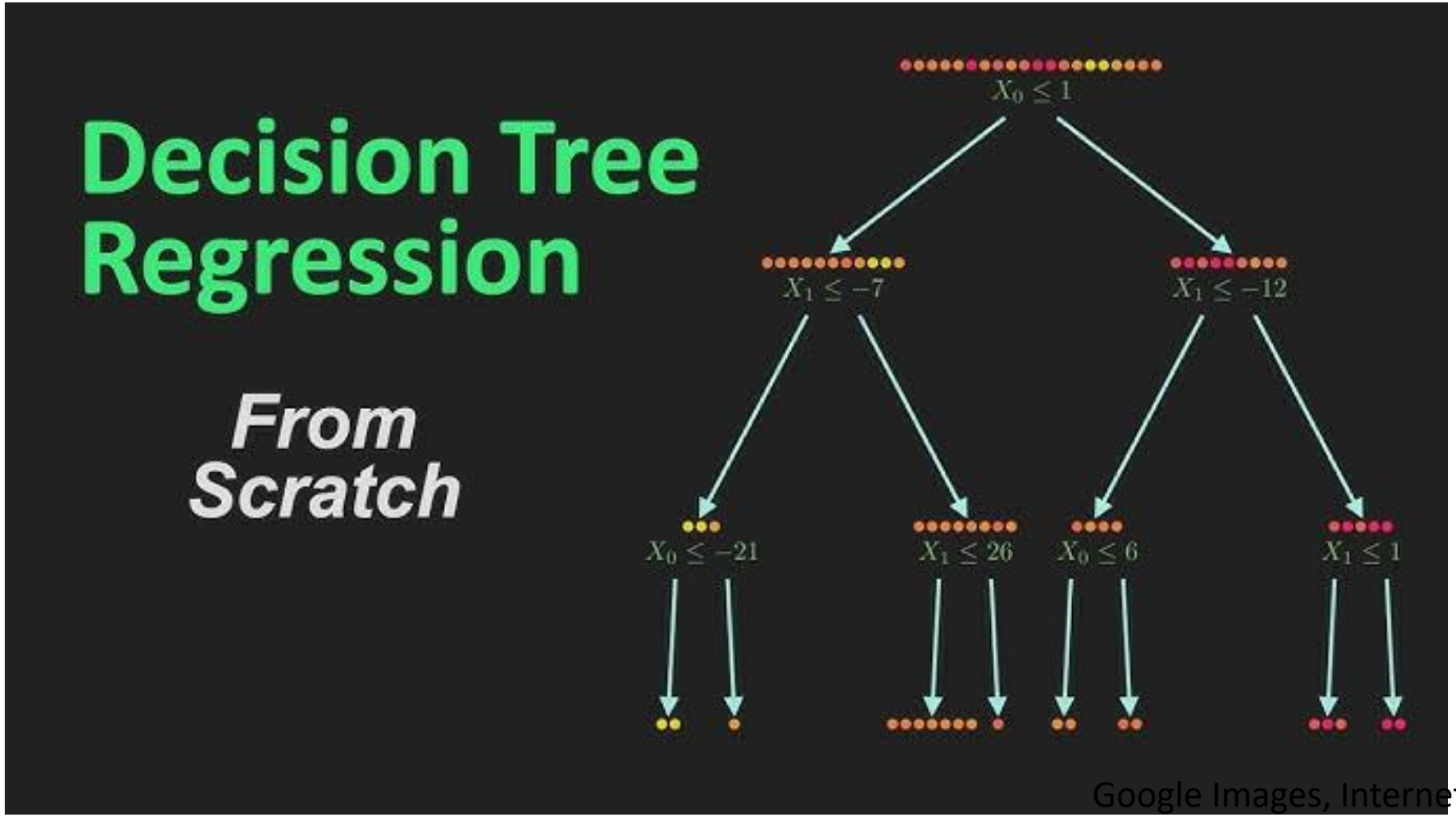


Decision Tree and Regression



Google Images, Internet

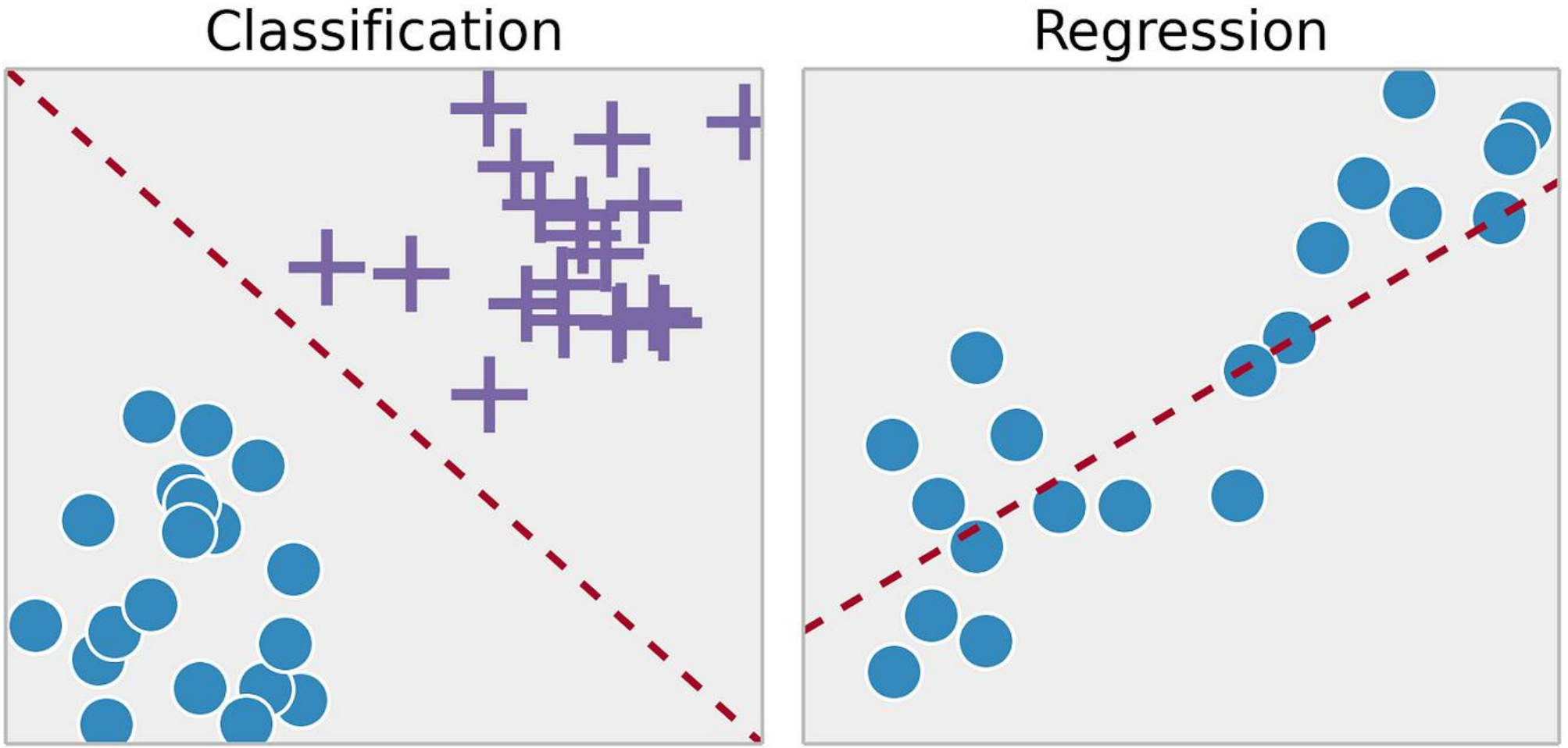
Regression and Decision Trees



Google Images, Internet

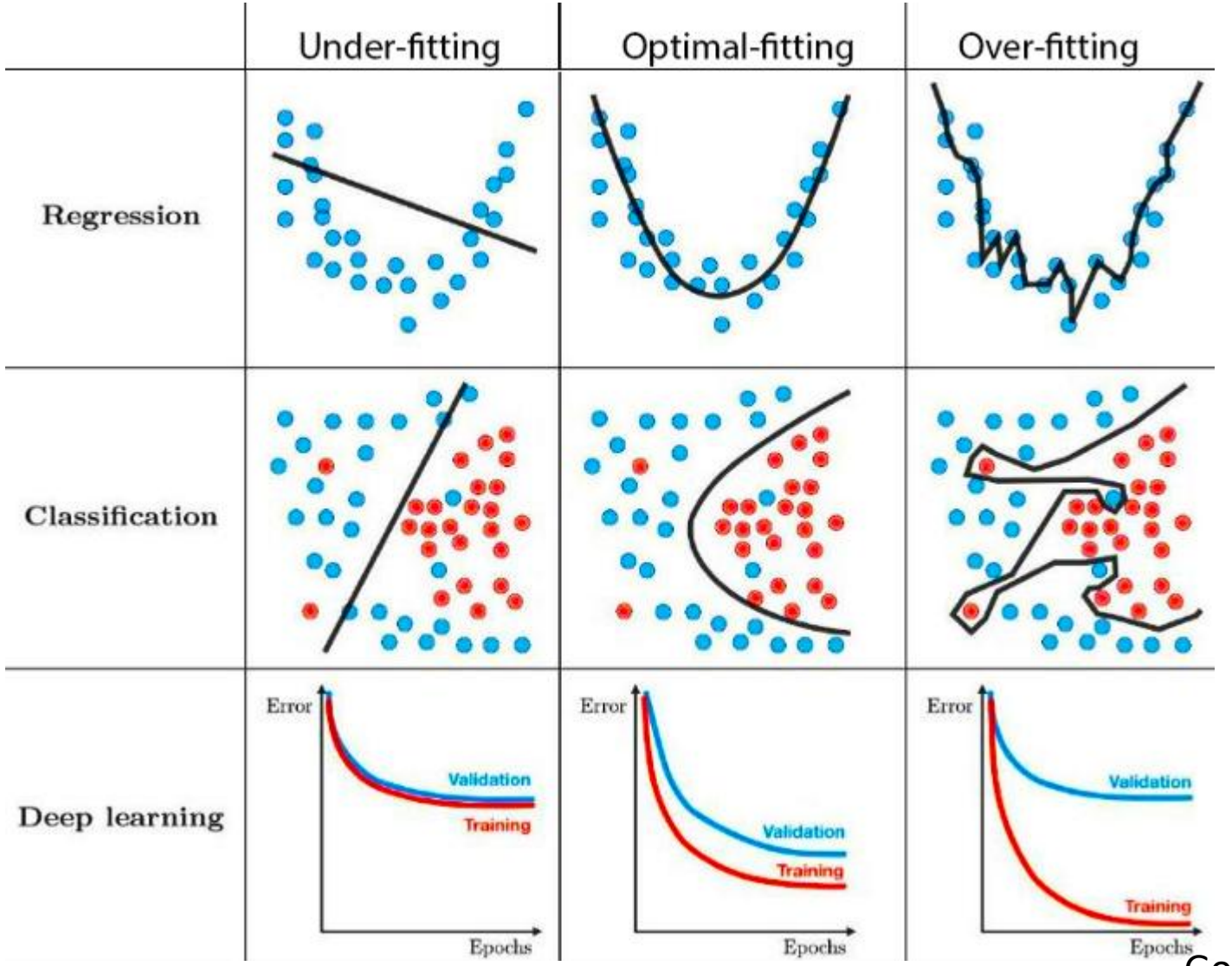
Google Images, Internet

Supervised Learning



Google Images, Internet

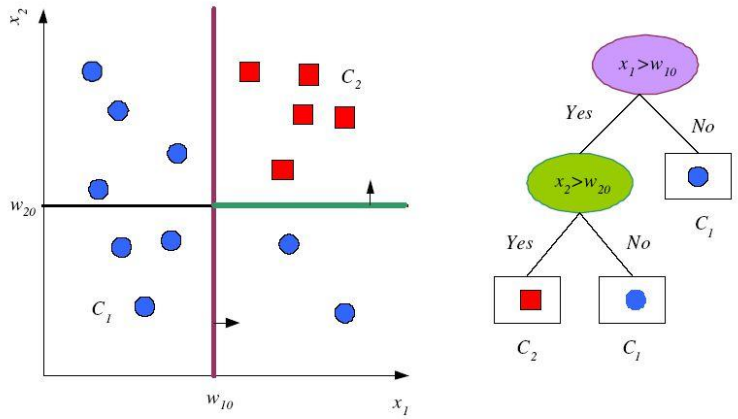
Underfit, Overfit, Optimal Fit

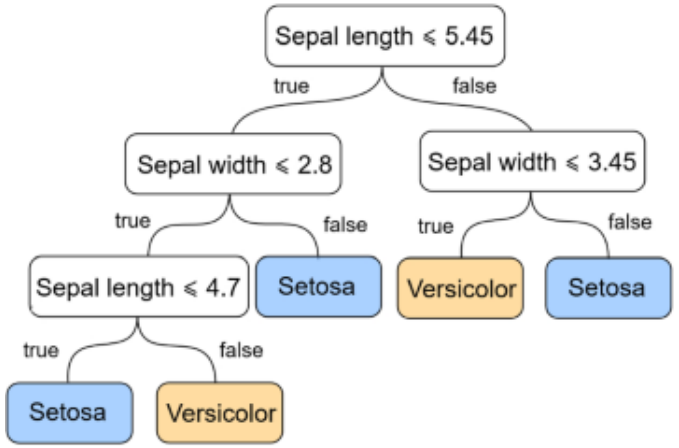


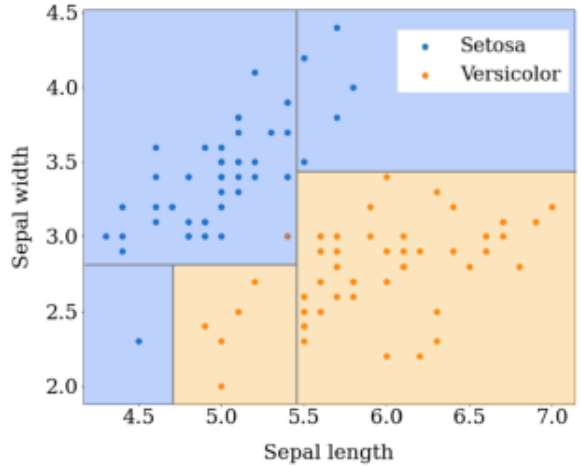
Google Images, Internet

Decision Tree Approaches

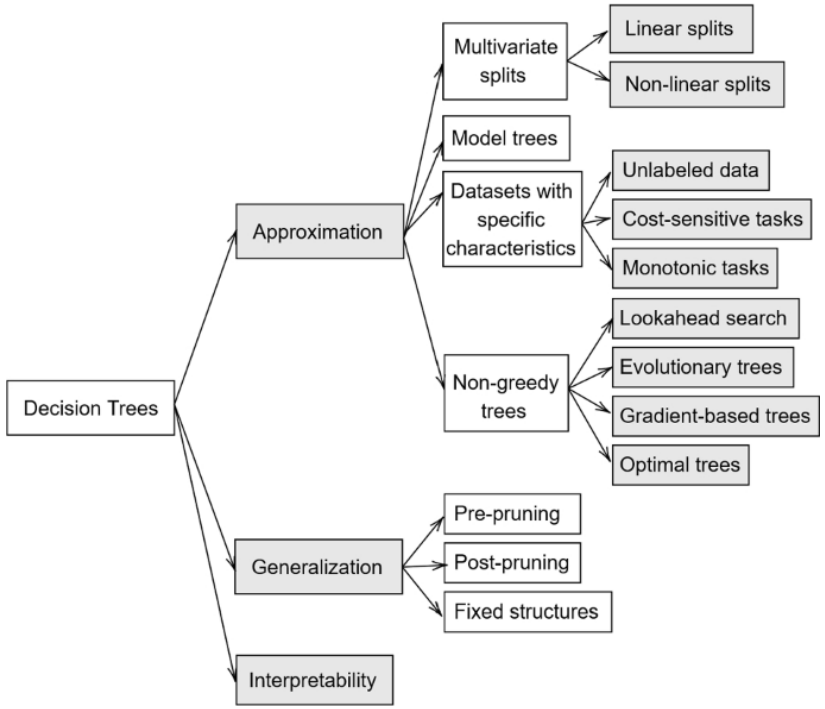
Tree Uses Nodes, and Leaves





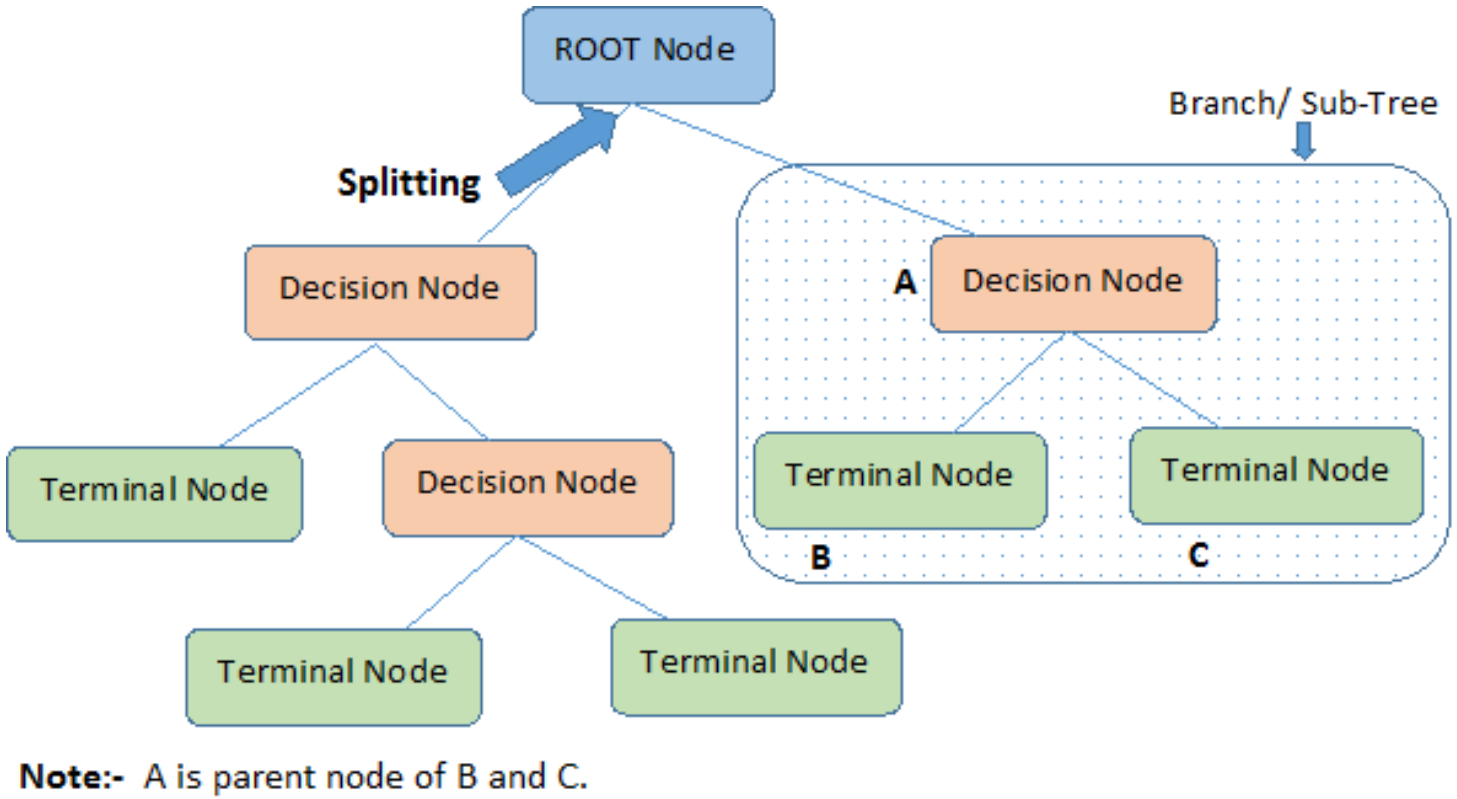


(b) Partitioning visualization

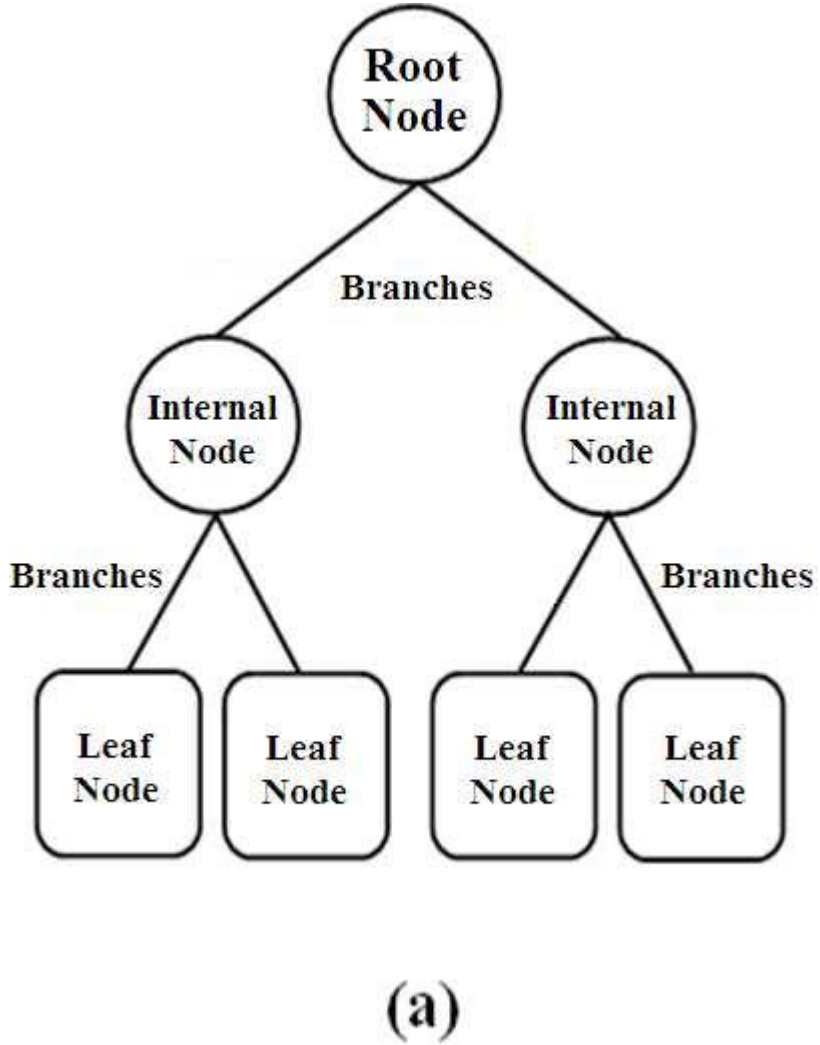


Google Images, Internet

Decision Tree, Node Types

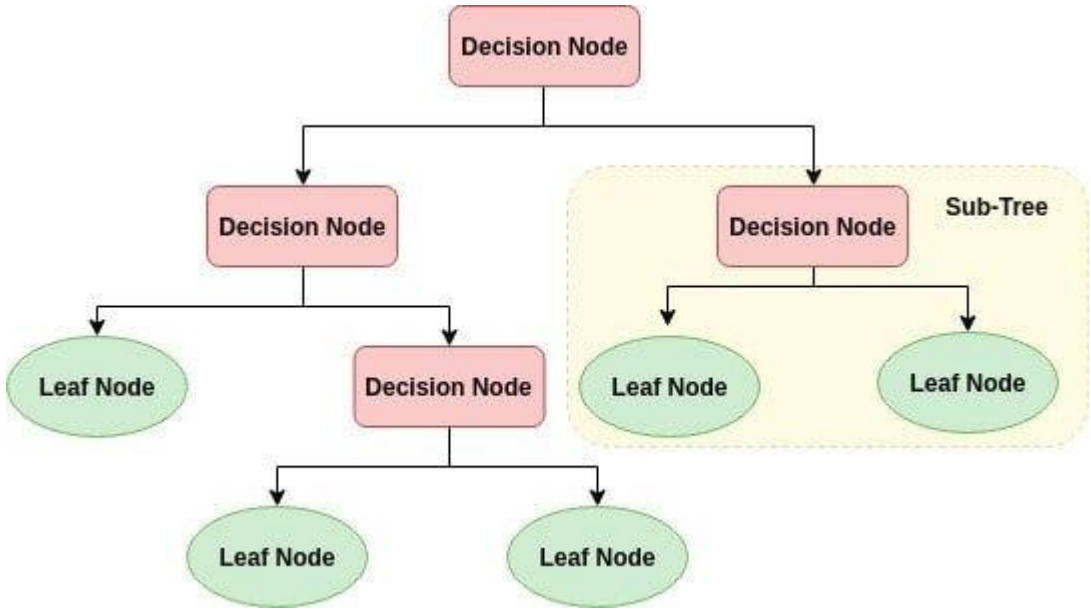


Note:- A is parent node of B and C.



Google Images, Internet

Node Types



Google Images, Internet

Pros and Cons: Decision Trees

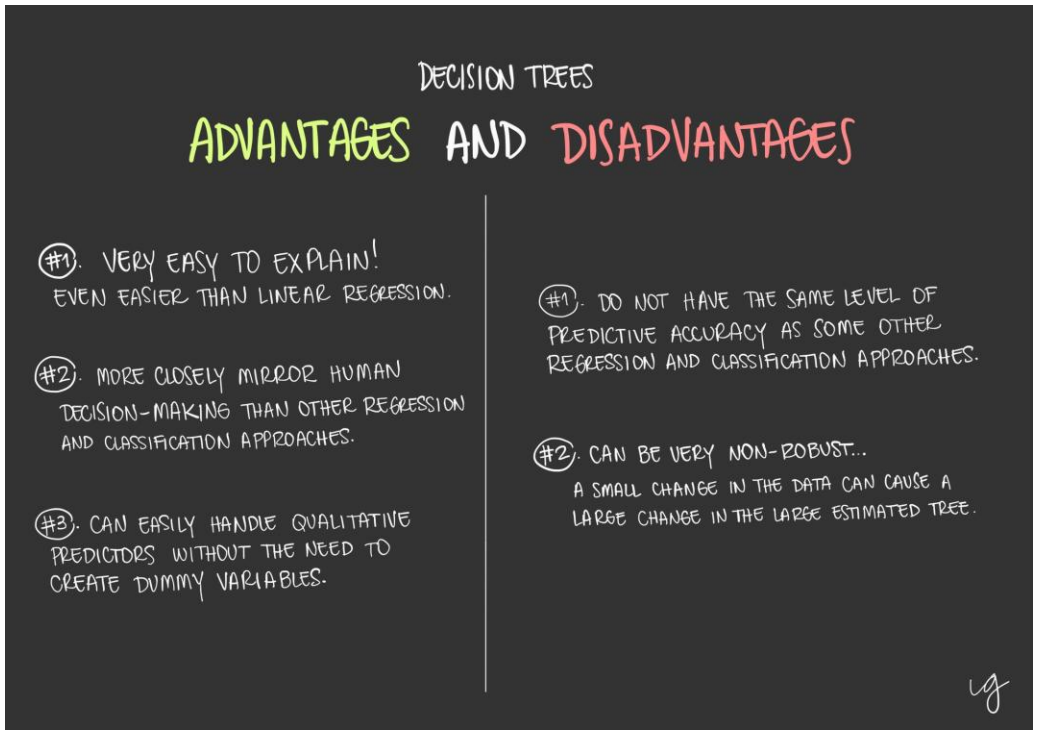
Decision Tree Pros and Cons

Pros

- No data assumptions
- Non-linear
- Discontinuous

Cons

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



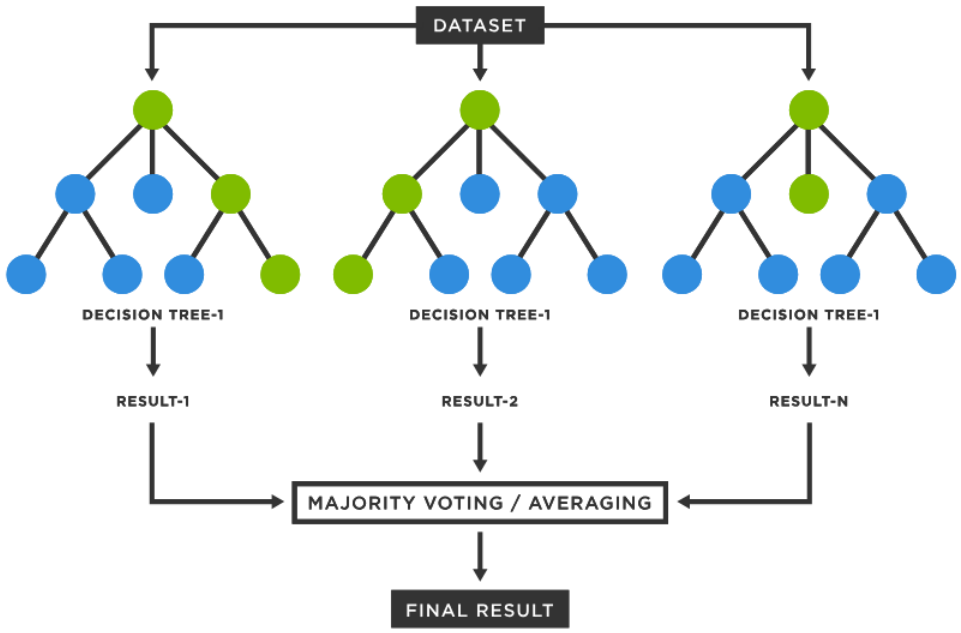
Google Images, Internet

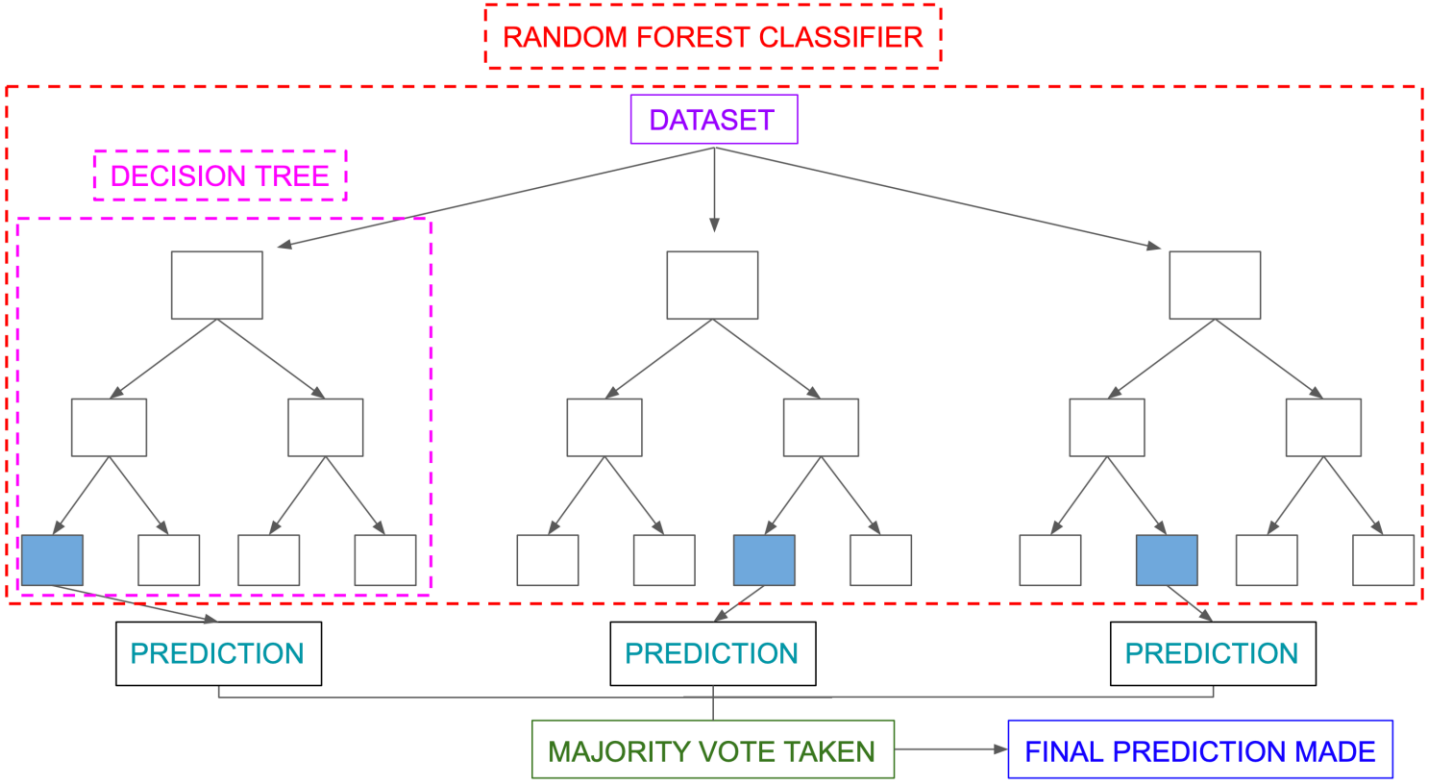
Decision Trees vs Random Forest

	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)

Google Images, Internet

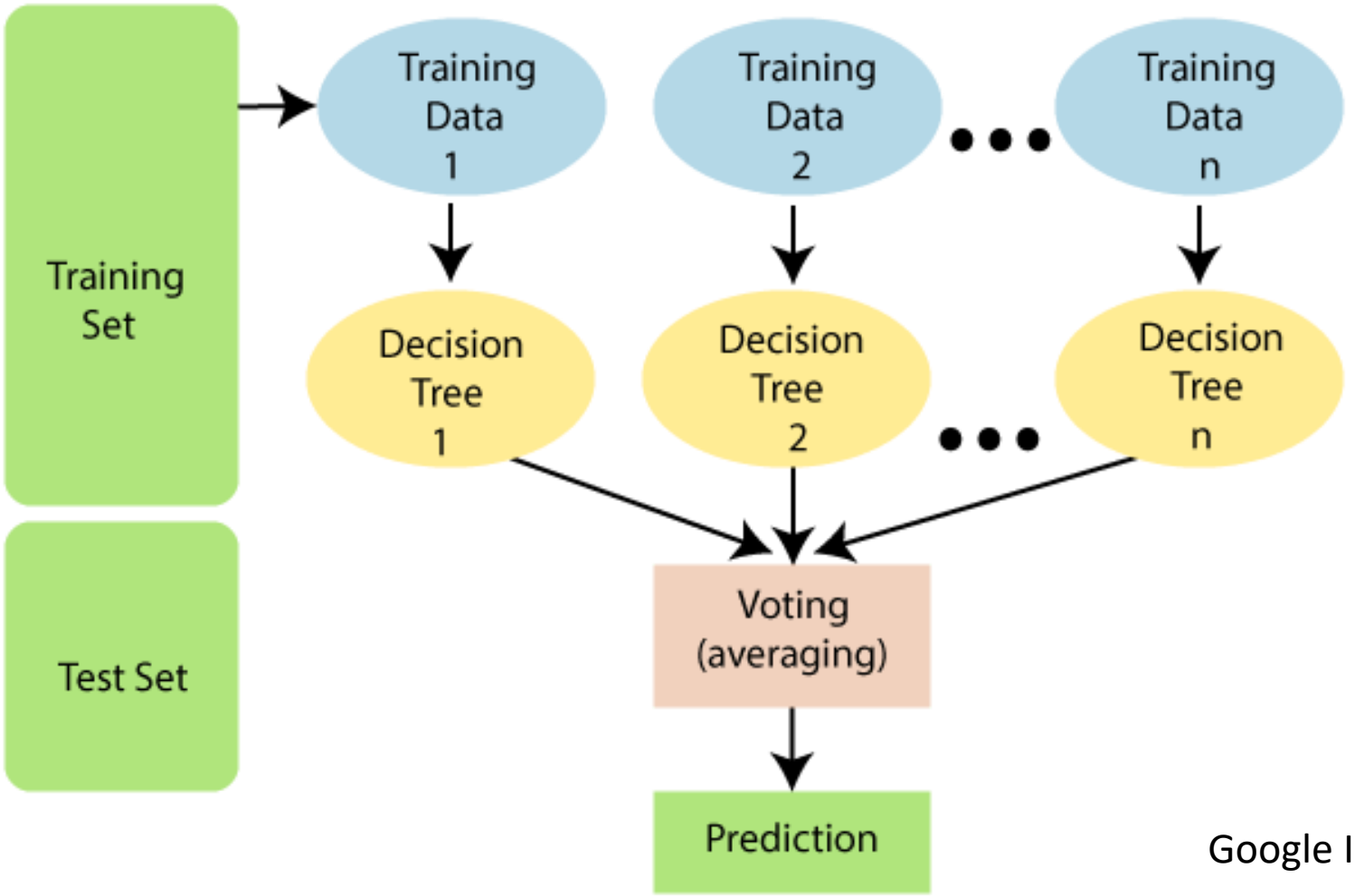
Random Forest





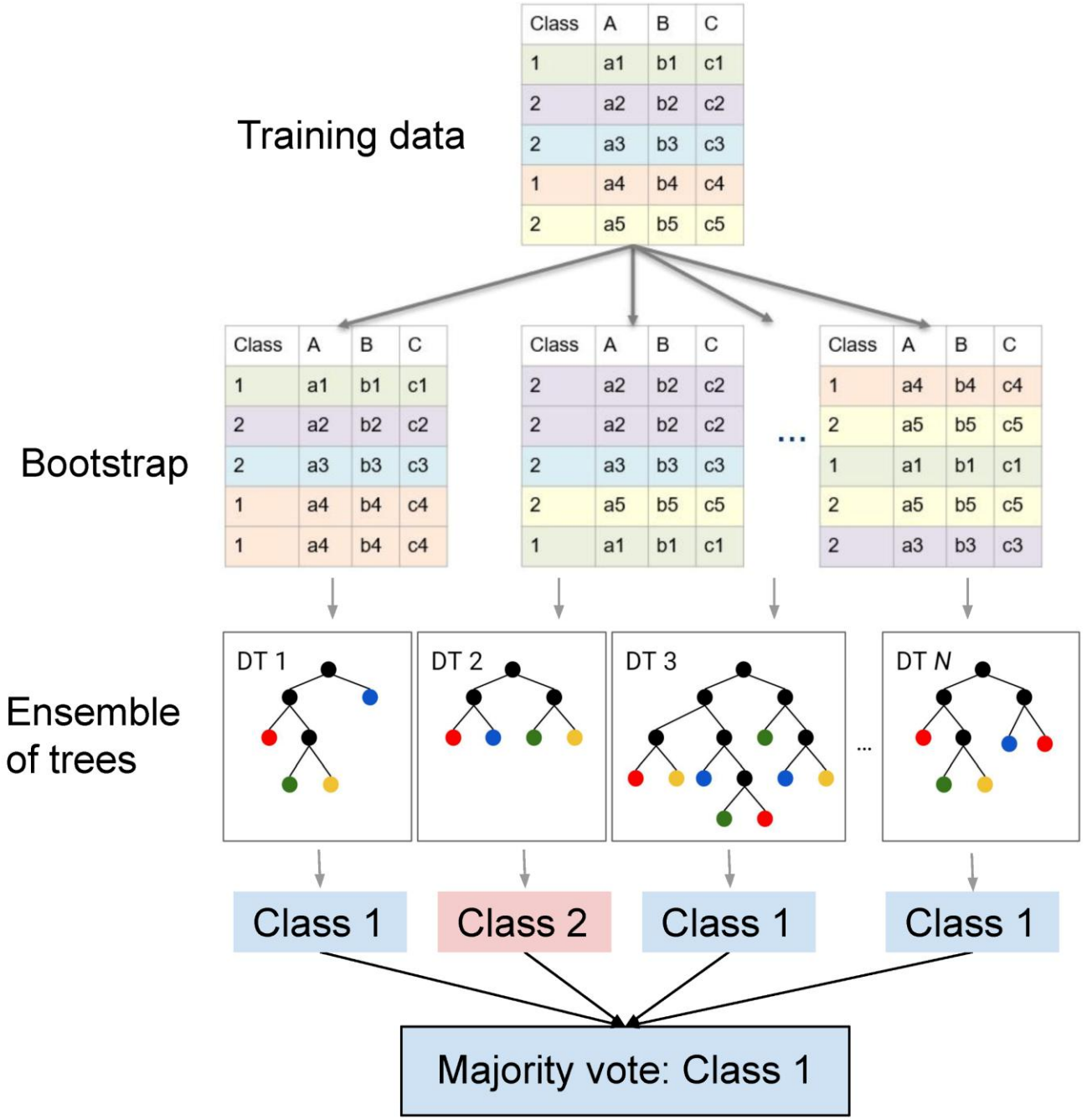
Google Images, Internet

Random Forest



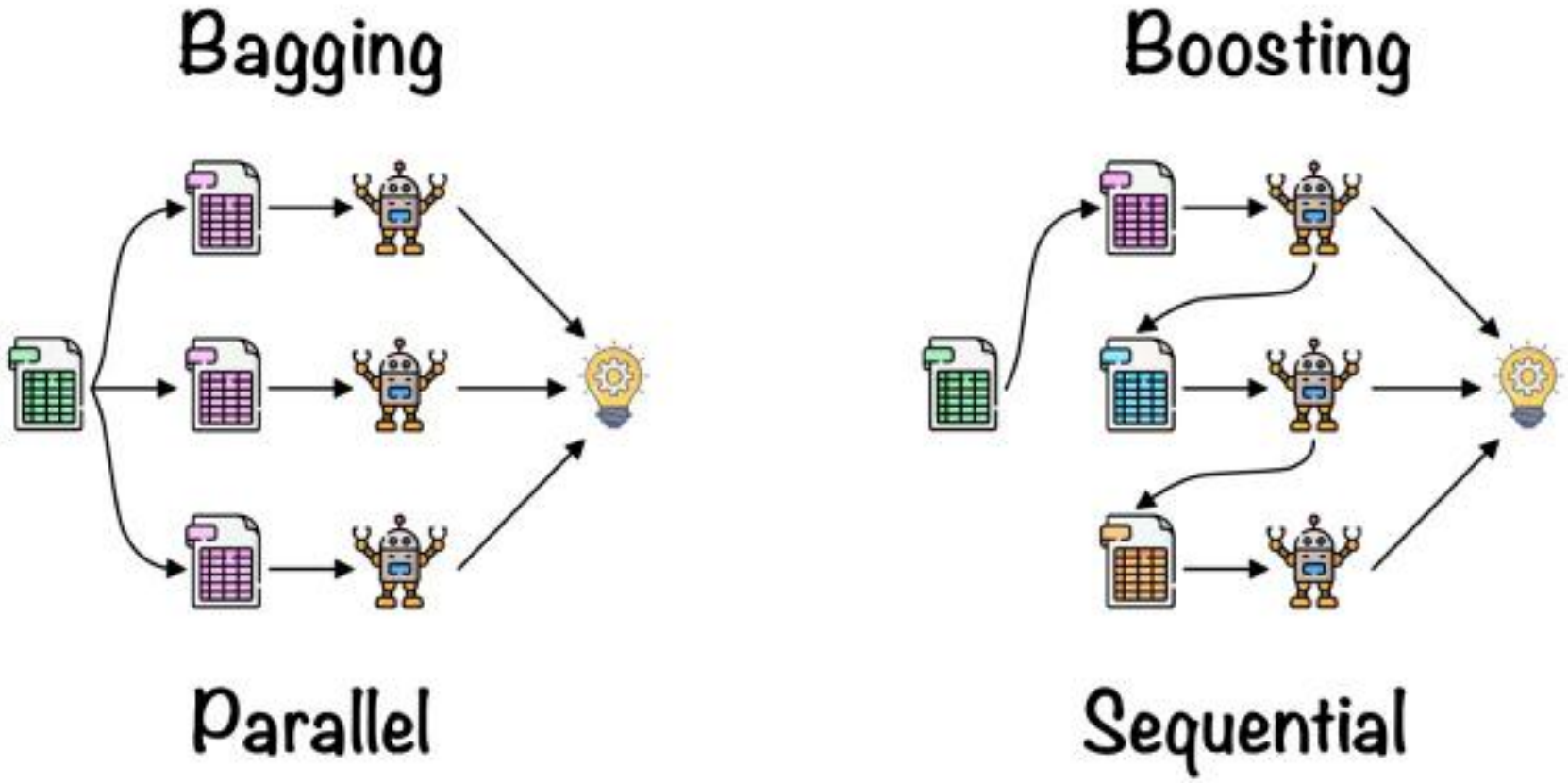
Google Images, Internet

Random Forest



Google Images, Internet

Random Forest Algorithms



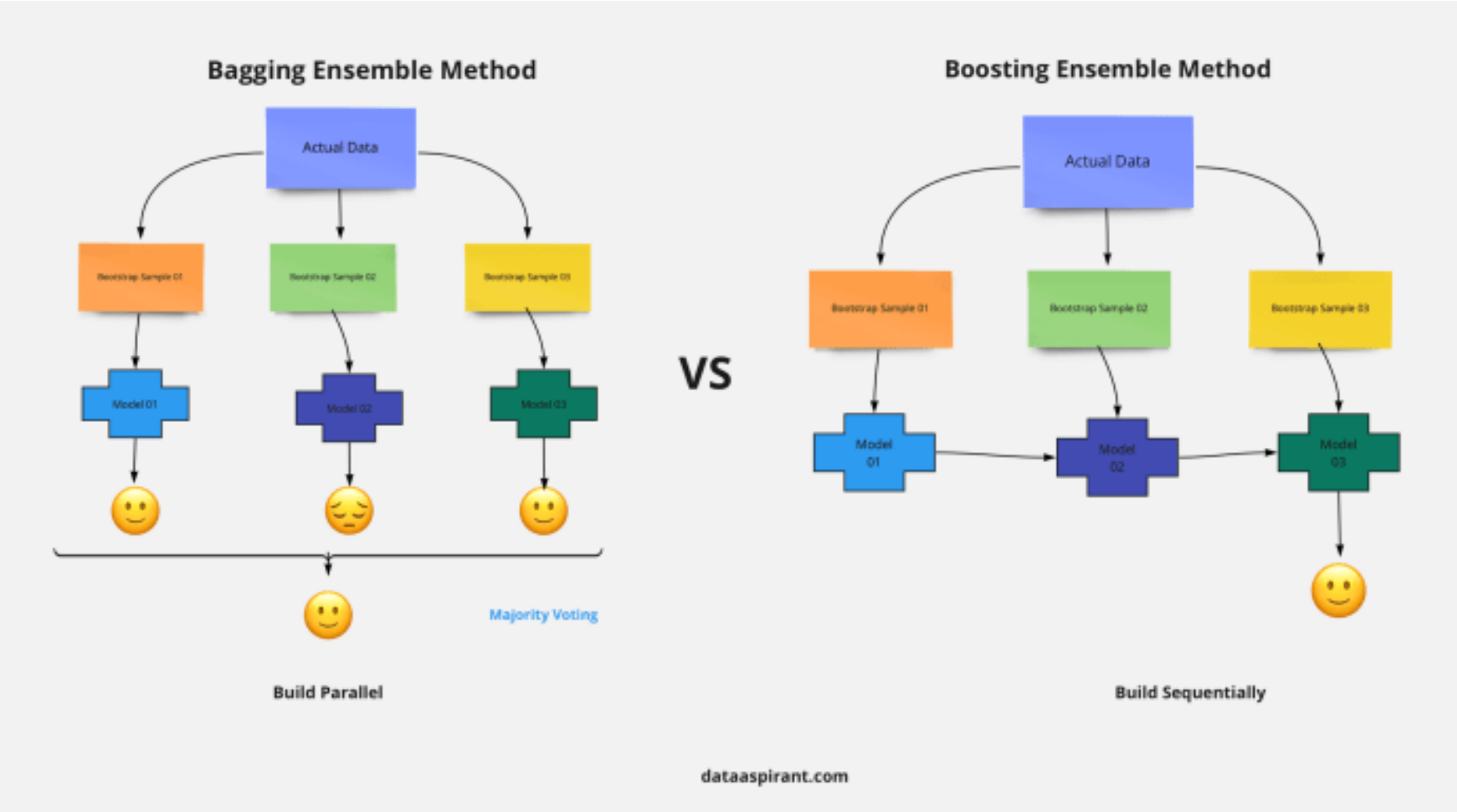
Google Images, Internet

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

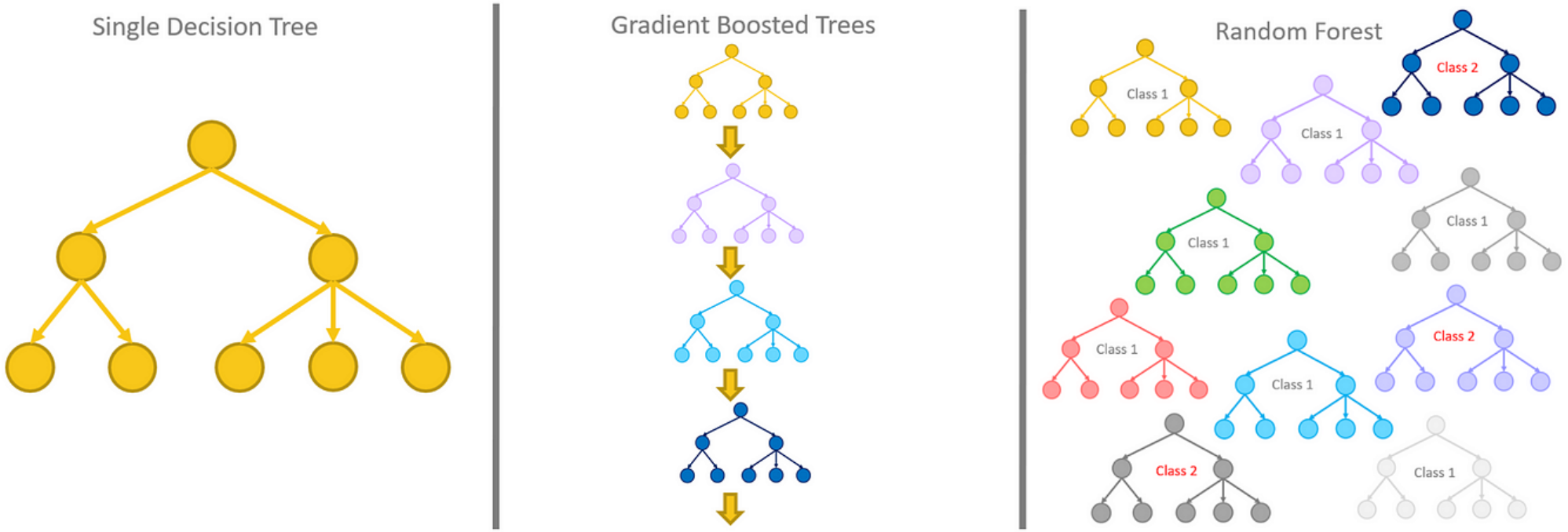
Google Images, Internet

Ensemble Methods



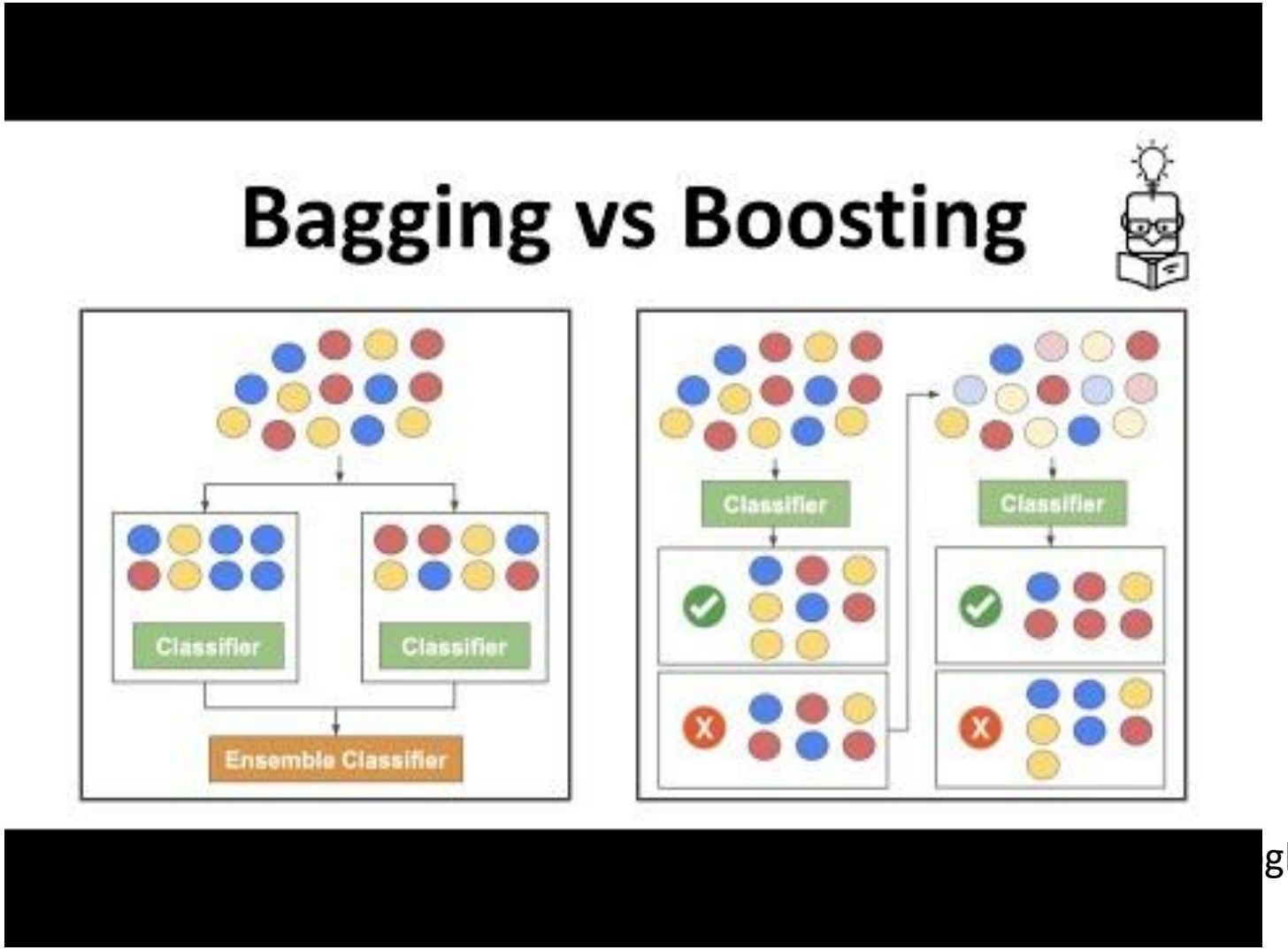
Google Images, Internet

Gradient Boosted Trees



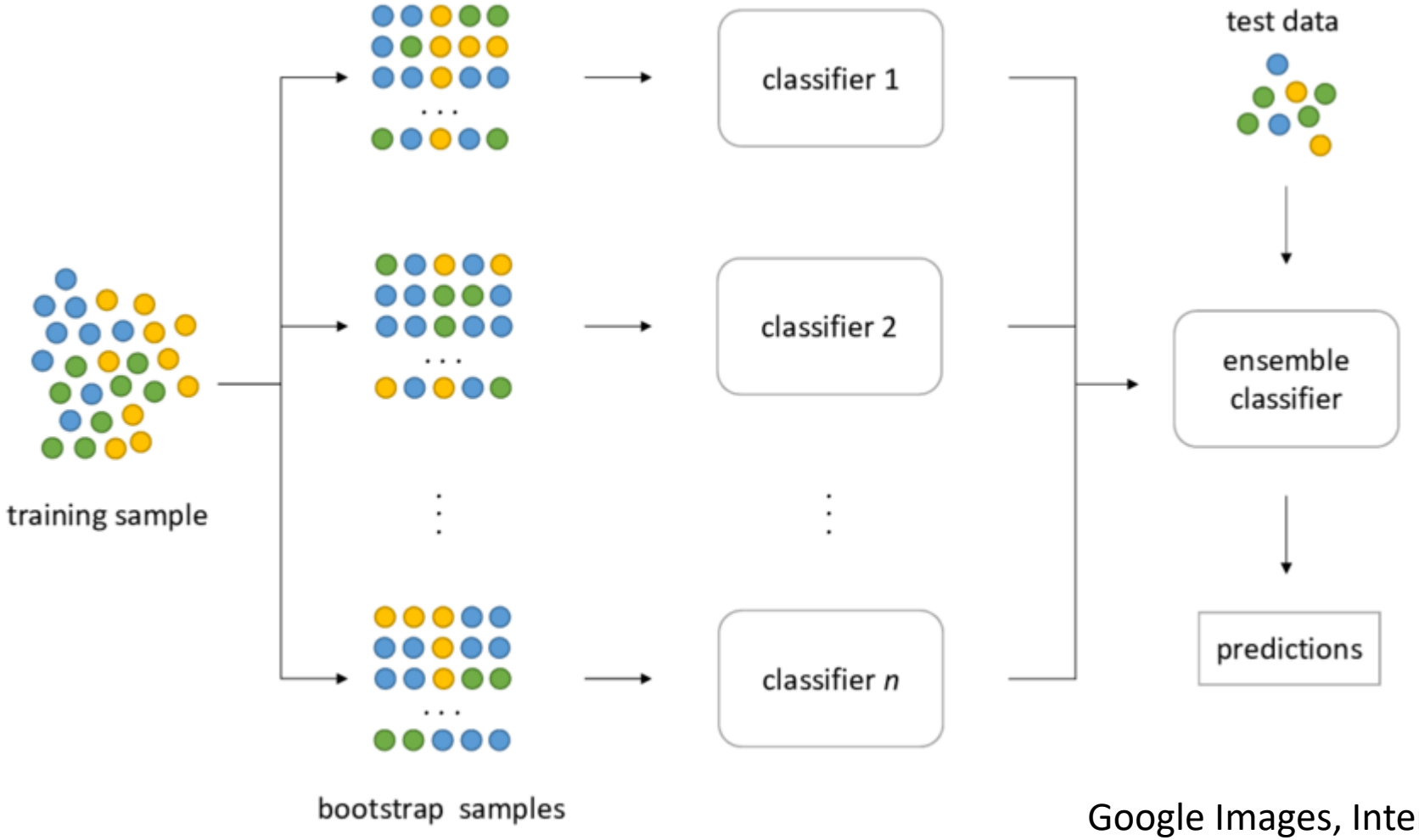
Google Images, Internet

Bagging vs Boosting



gle Images, Internet

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):

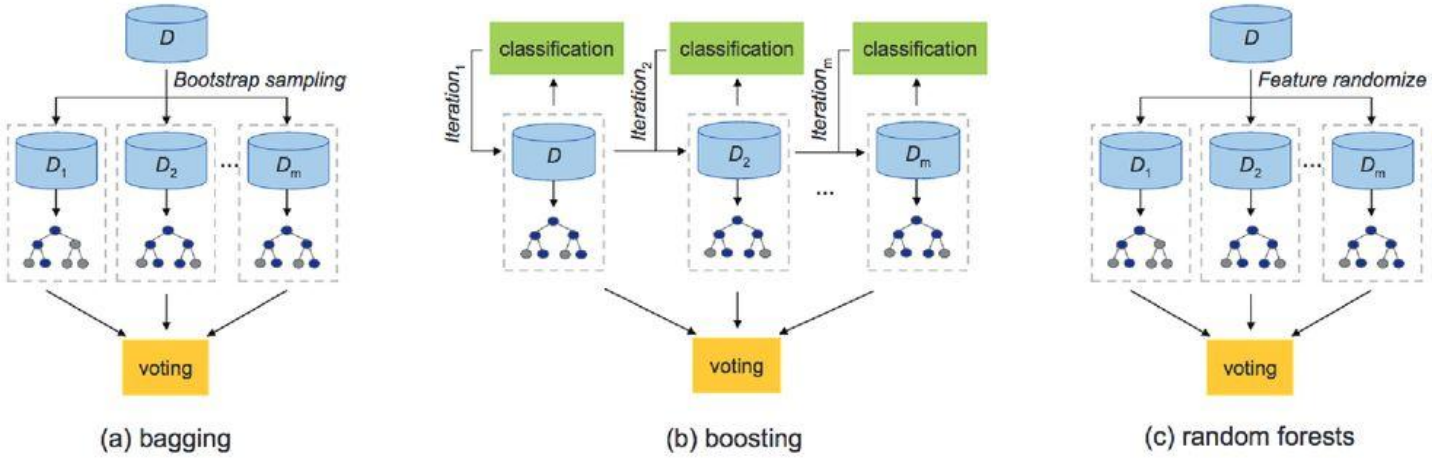
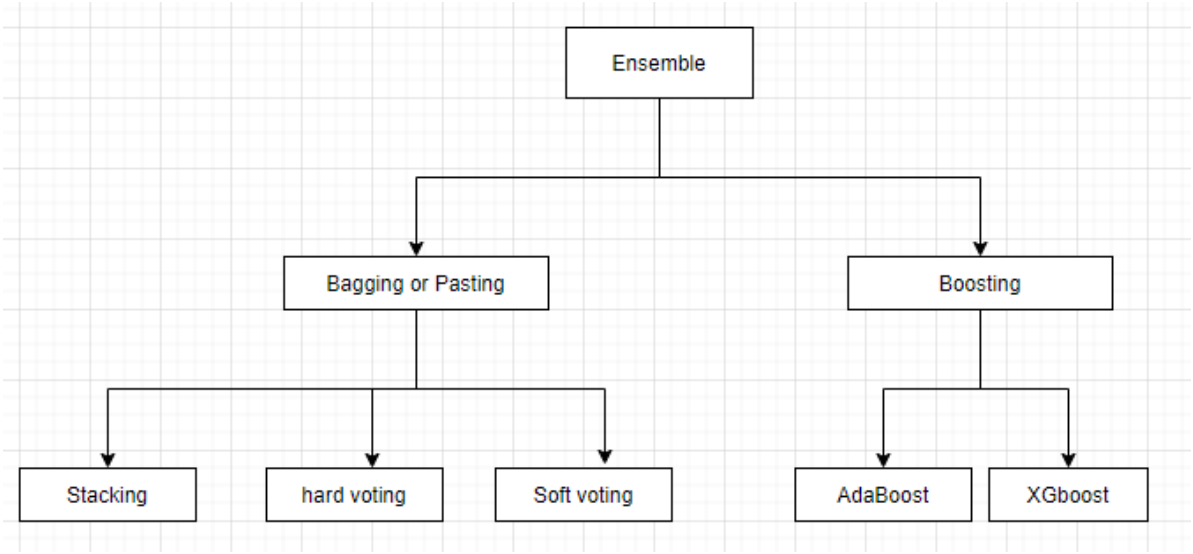


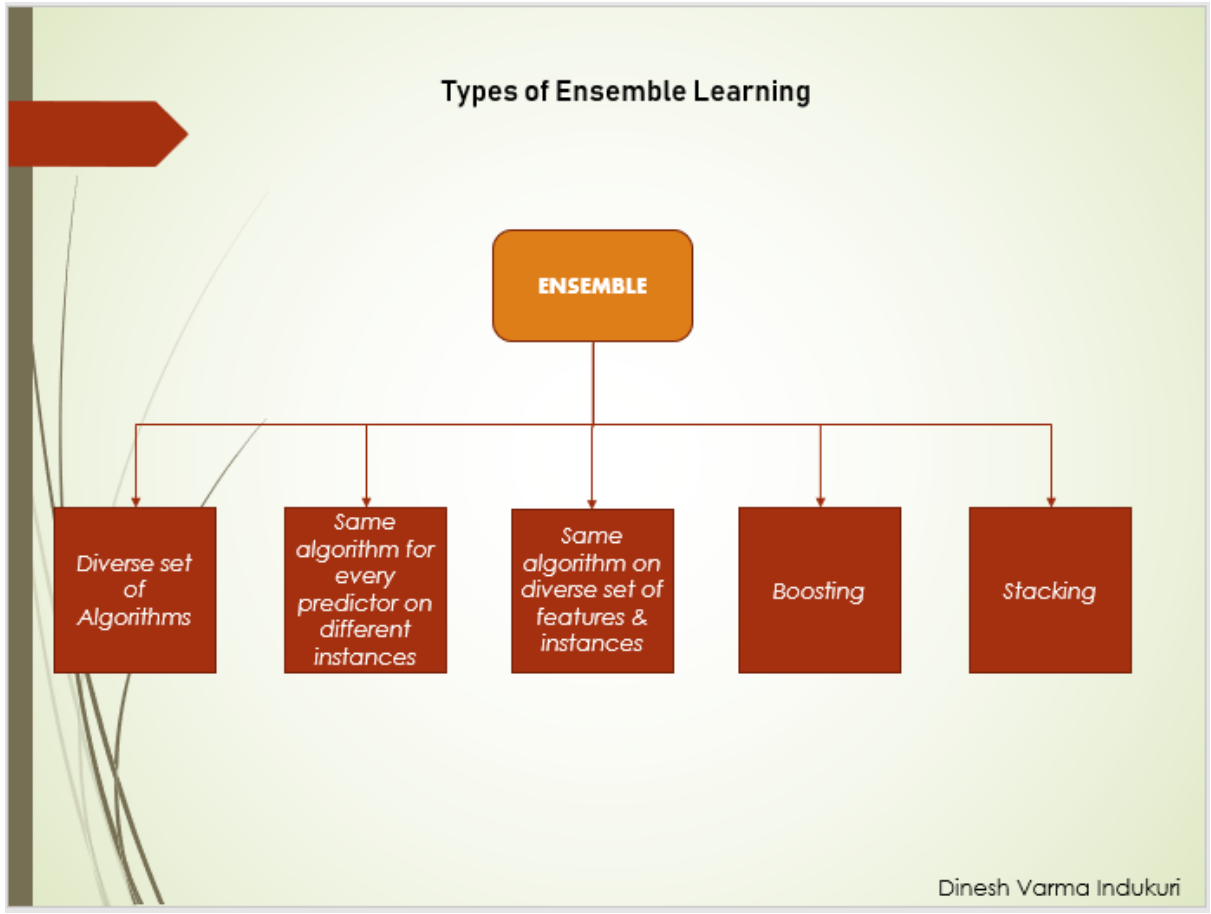
Fig. 1: Schematic illustration of the three popular ensemble methods.

Current Bioinformatics, 5, (4):296-308, 2010.

Google Images, Internet

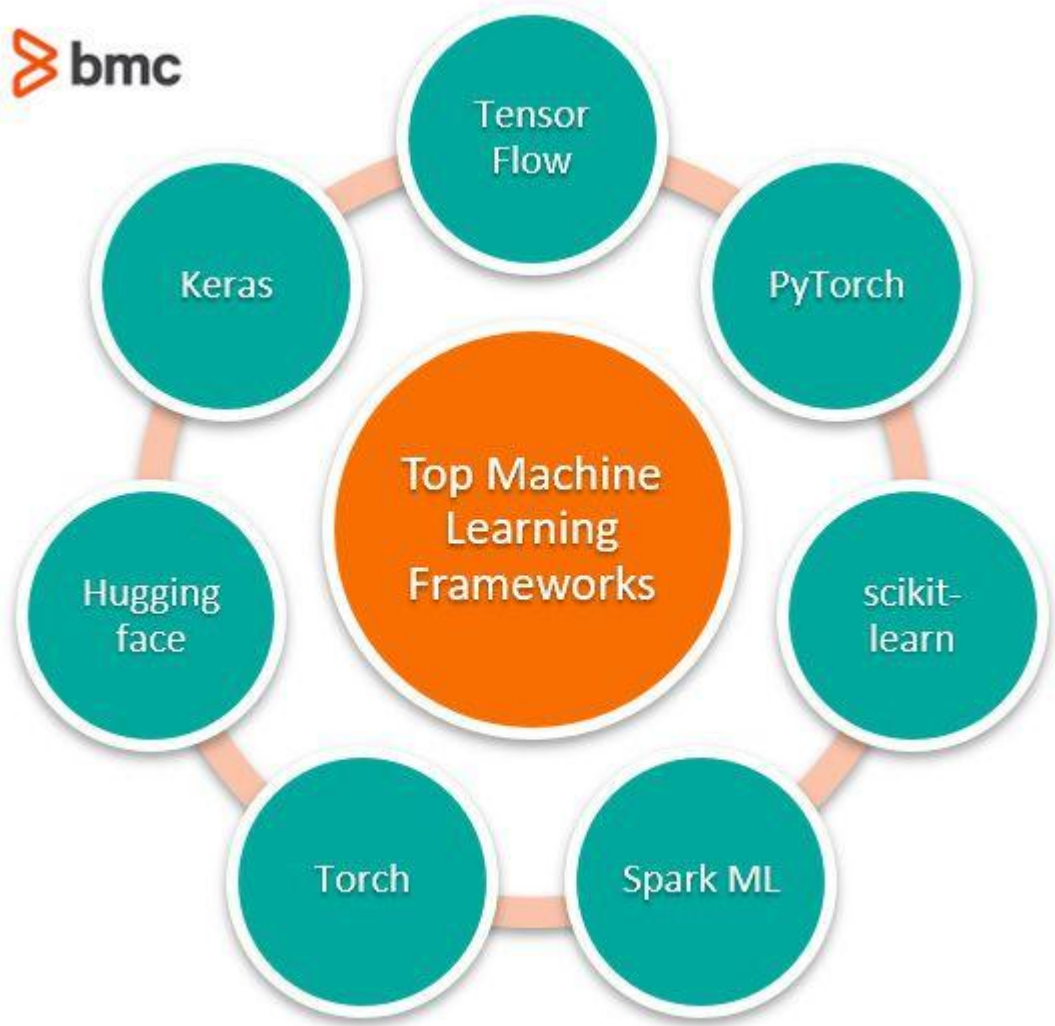
Types of Ensemble Learning

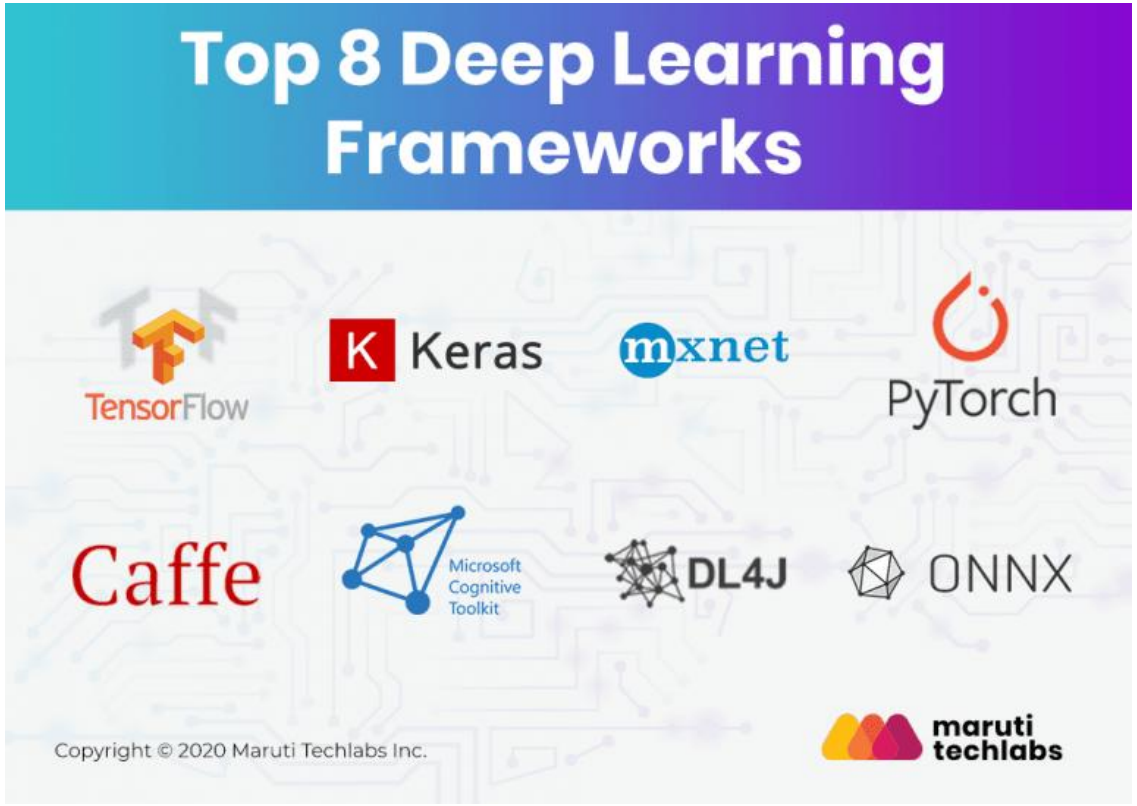




Google Images, Internet

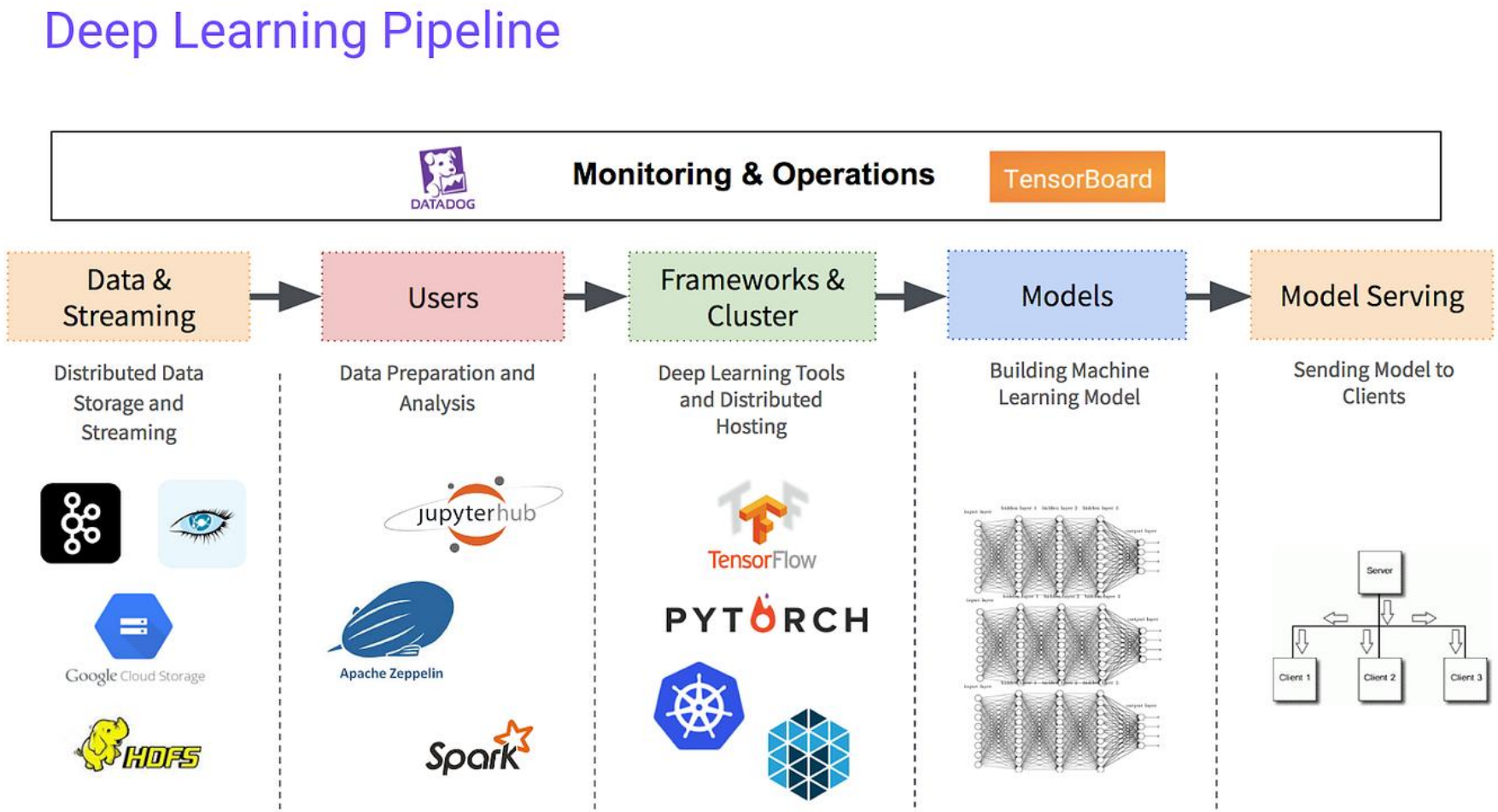
Machine Learning Frameworks





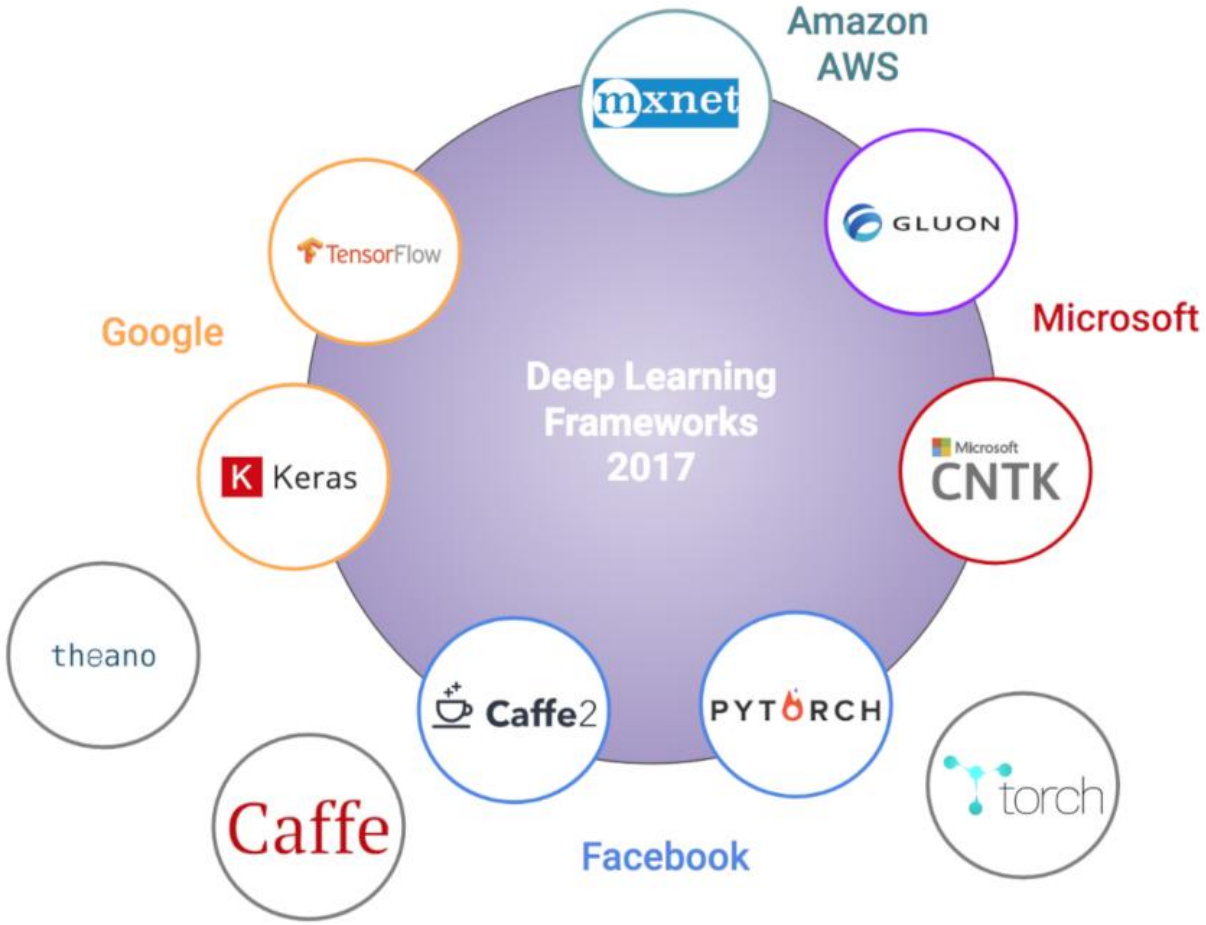
Google Images, Internet

Machine Learning Frameworks



Google Images, Internet

Machine Learning Frameworks



Google Images, Internet