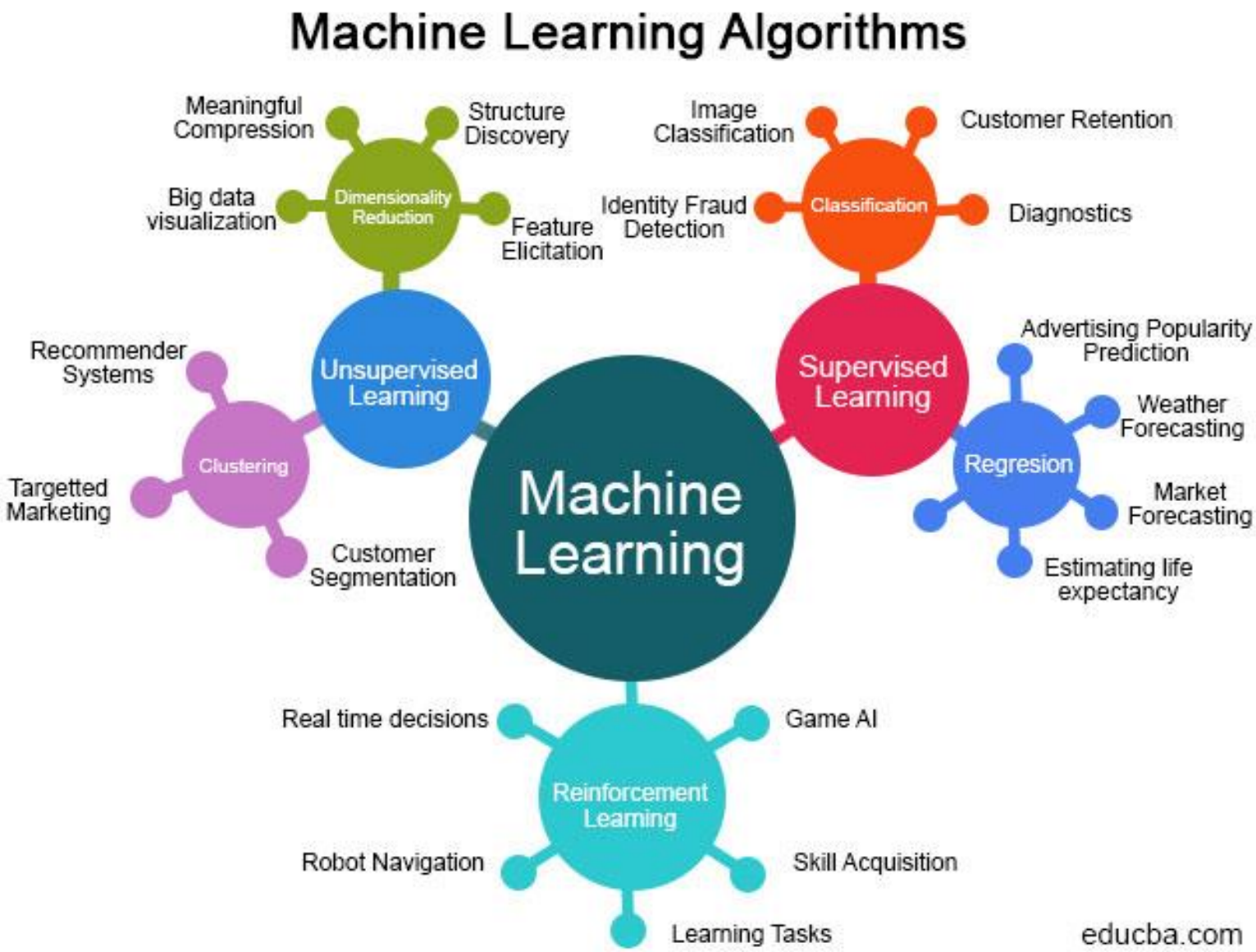
Random Forest Decision Model and Its Applications

Albert C. Yang, M.D., Ph.D.

Institutes of Brain Science/Digital Medicine Center
National Yang-Ming University

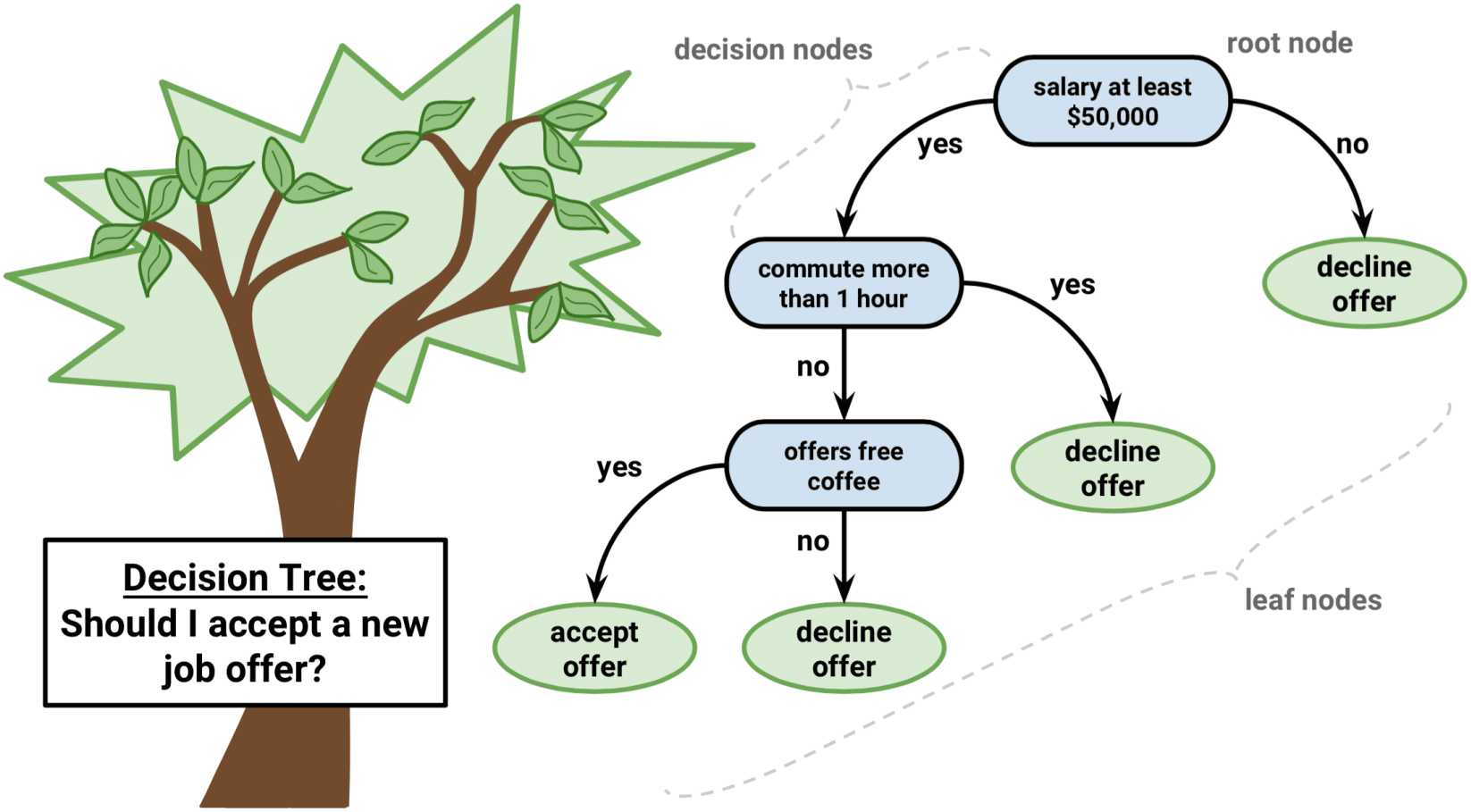
Apr 16, 2020

accyang@gmail.com



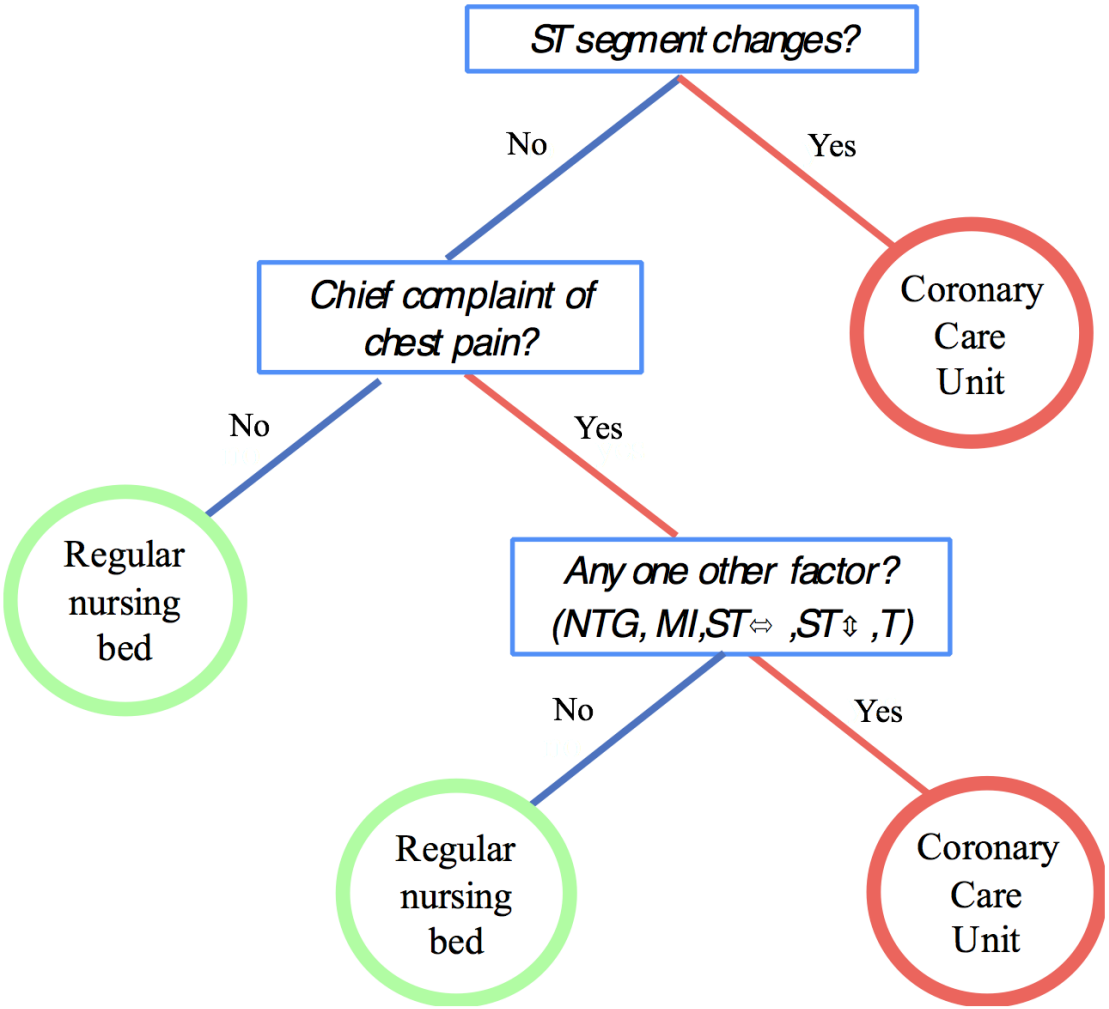
https://www.educba.com/machine-learning-algorithms/

Classification and Regression Tree



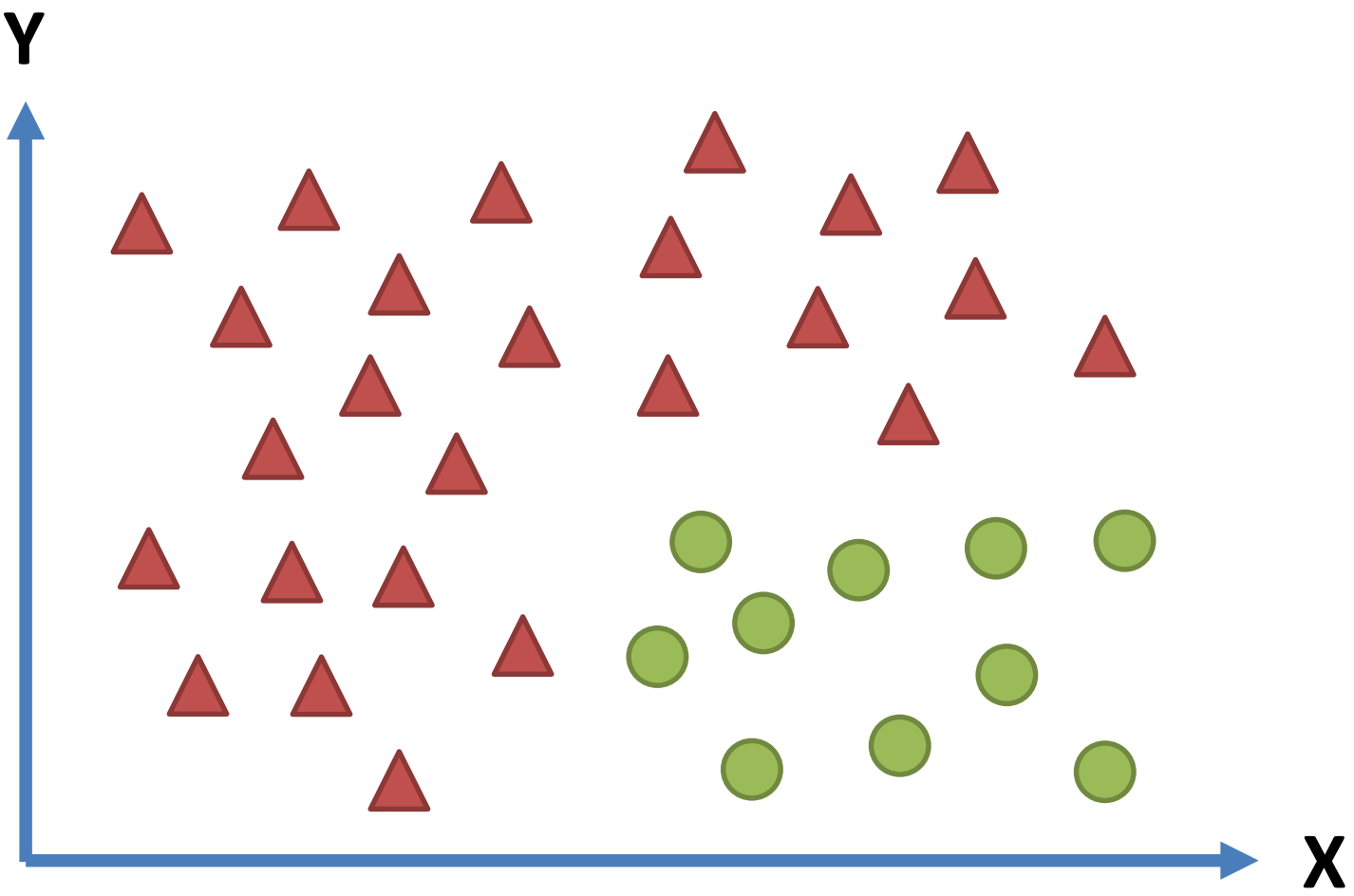
https://www.datacamp.com/community/tutorials/decision-trees-R

Decision Tree is Common in Medicine

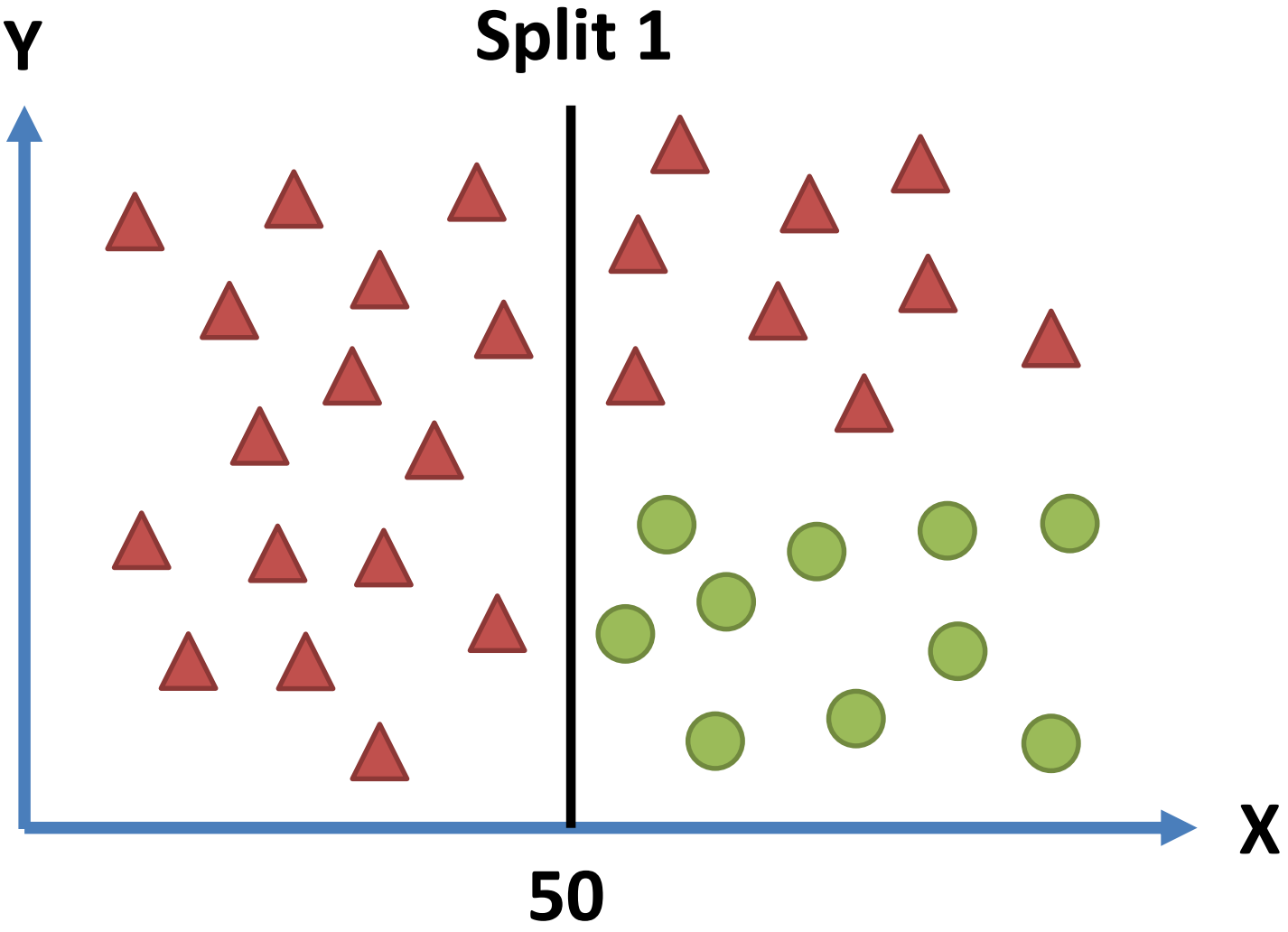


https://en.wikipedia.org/wiki/Fast-and-frugal_trees#/media/File:FFT-FIG01.png

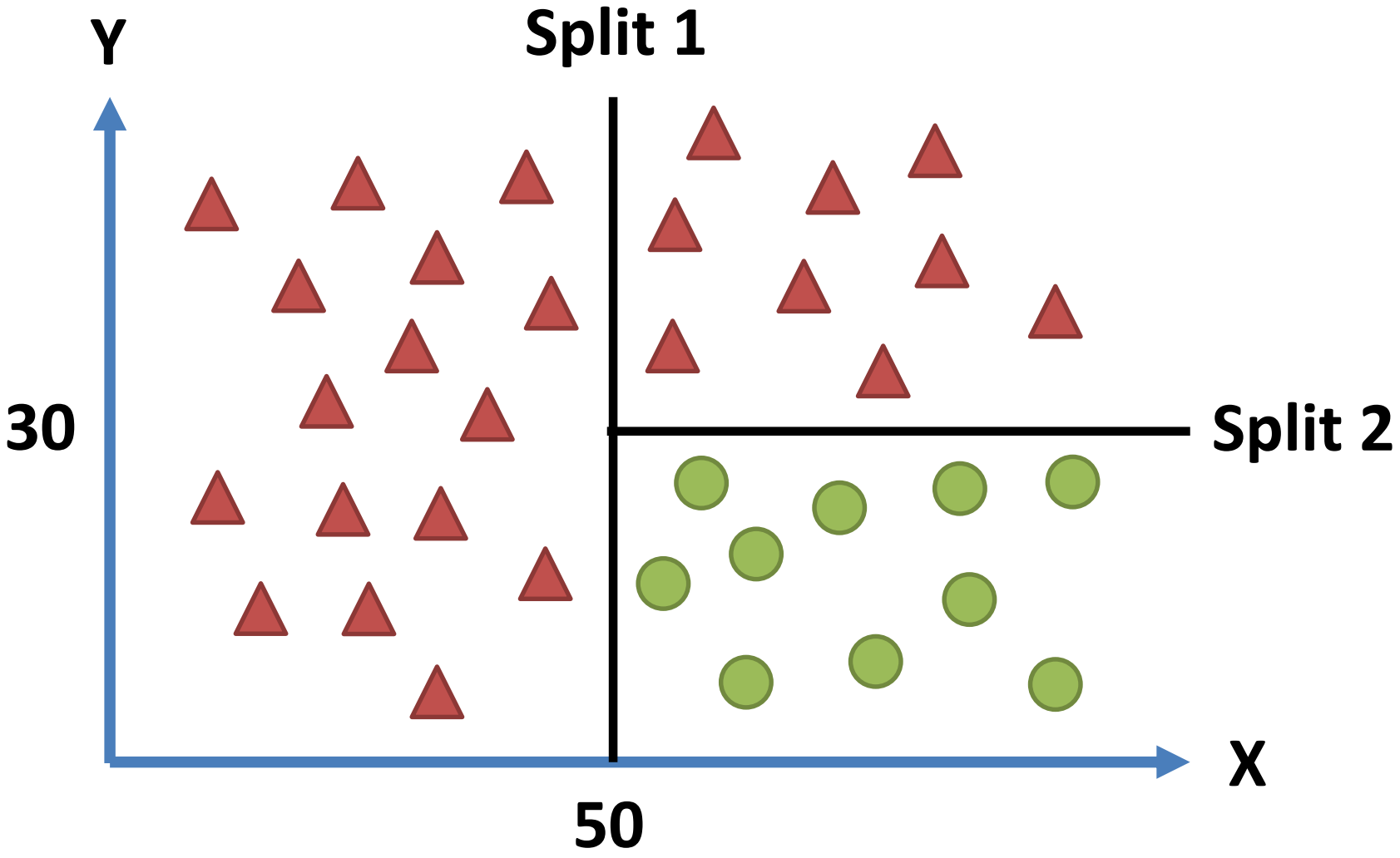
A Simple Classification Tree



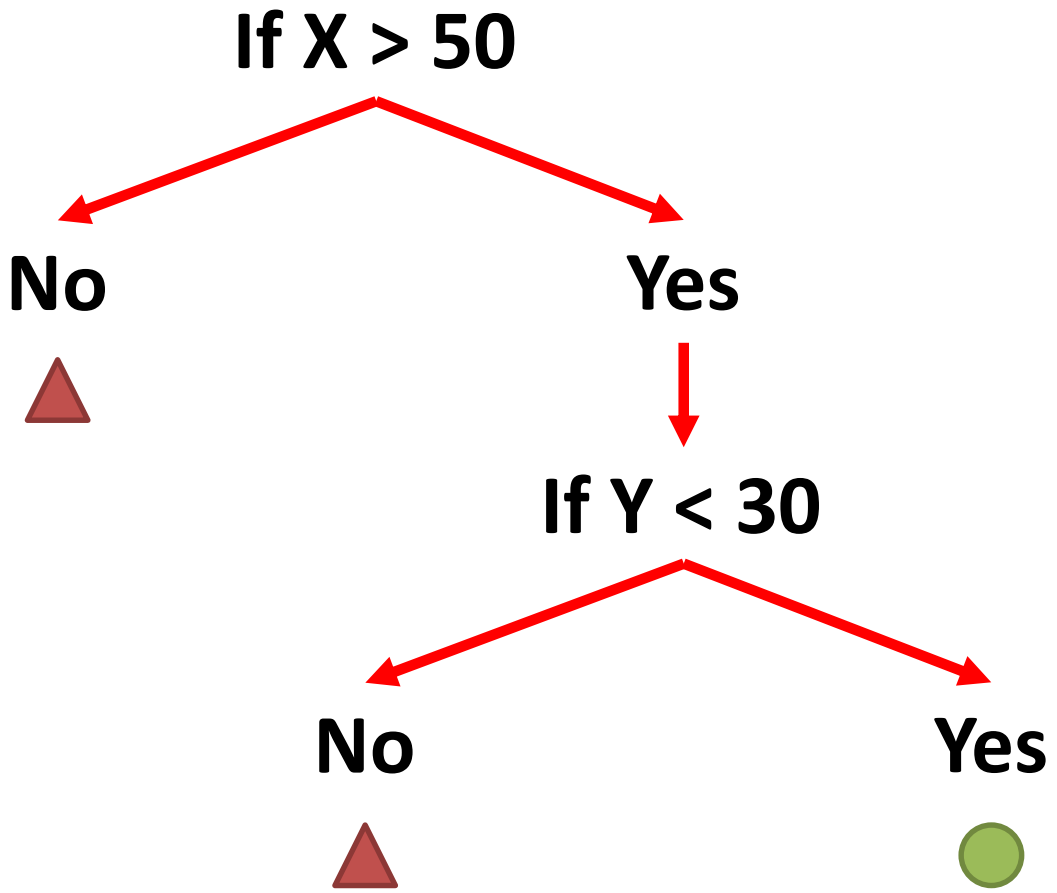
A Simple Classification Tree



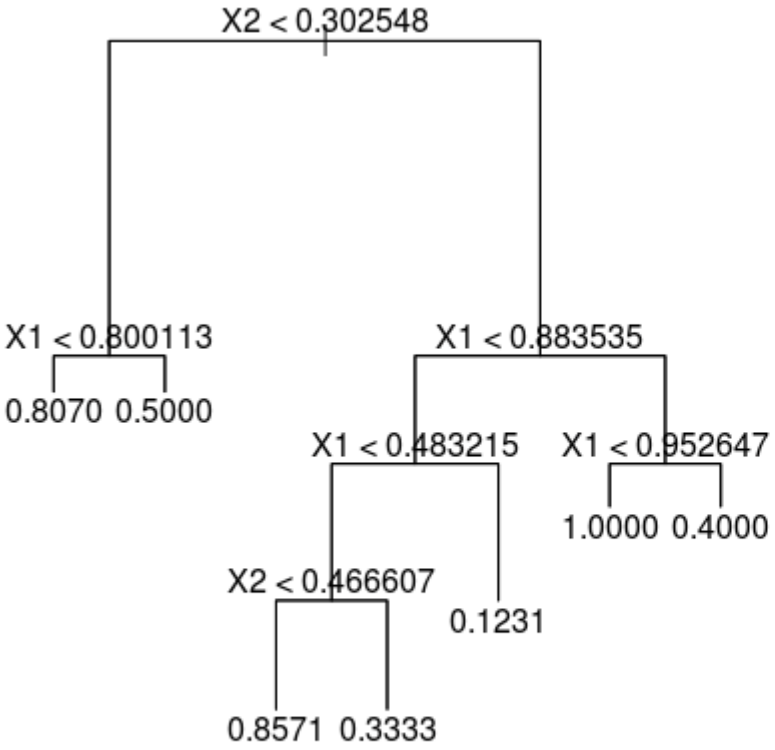
A Simple Classification Tree

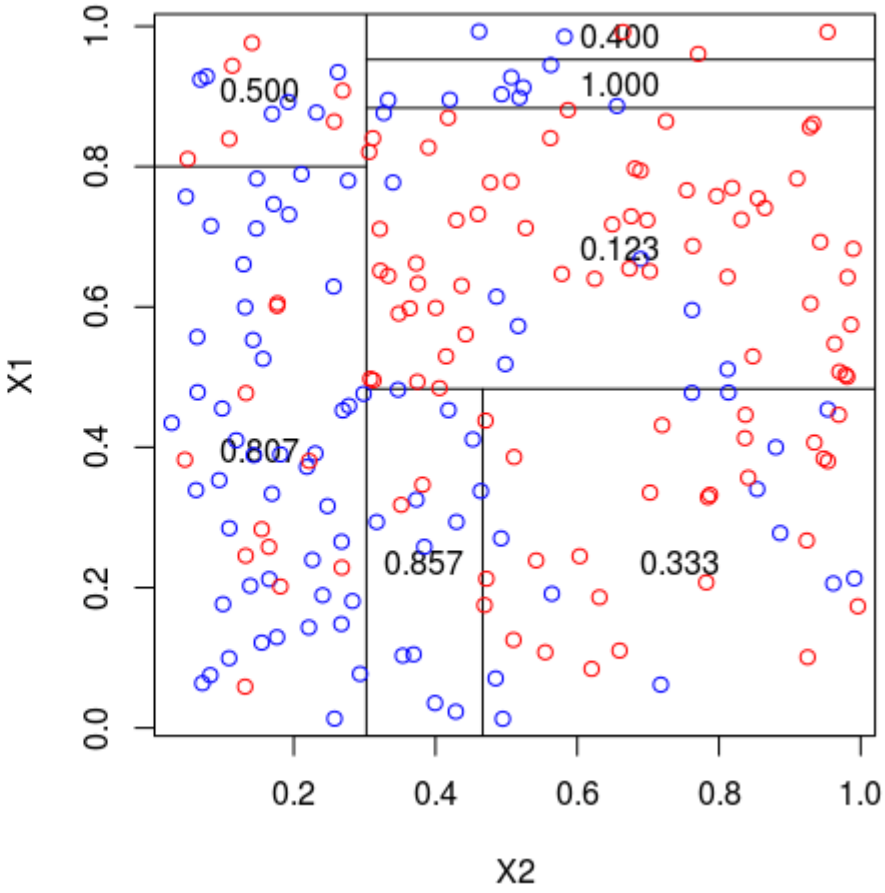


A Simple Classification Tree



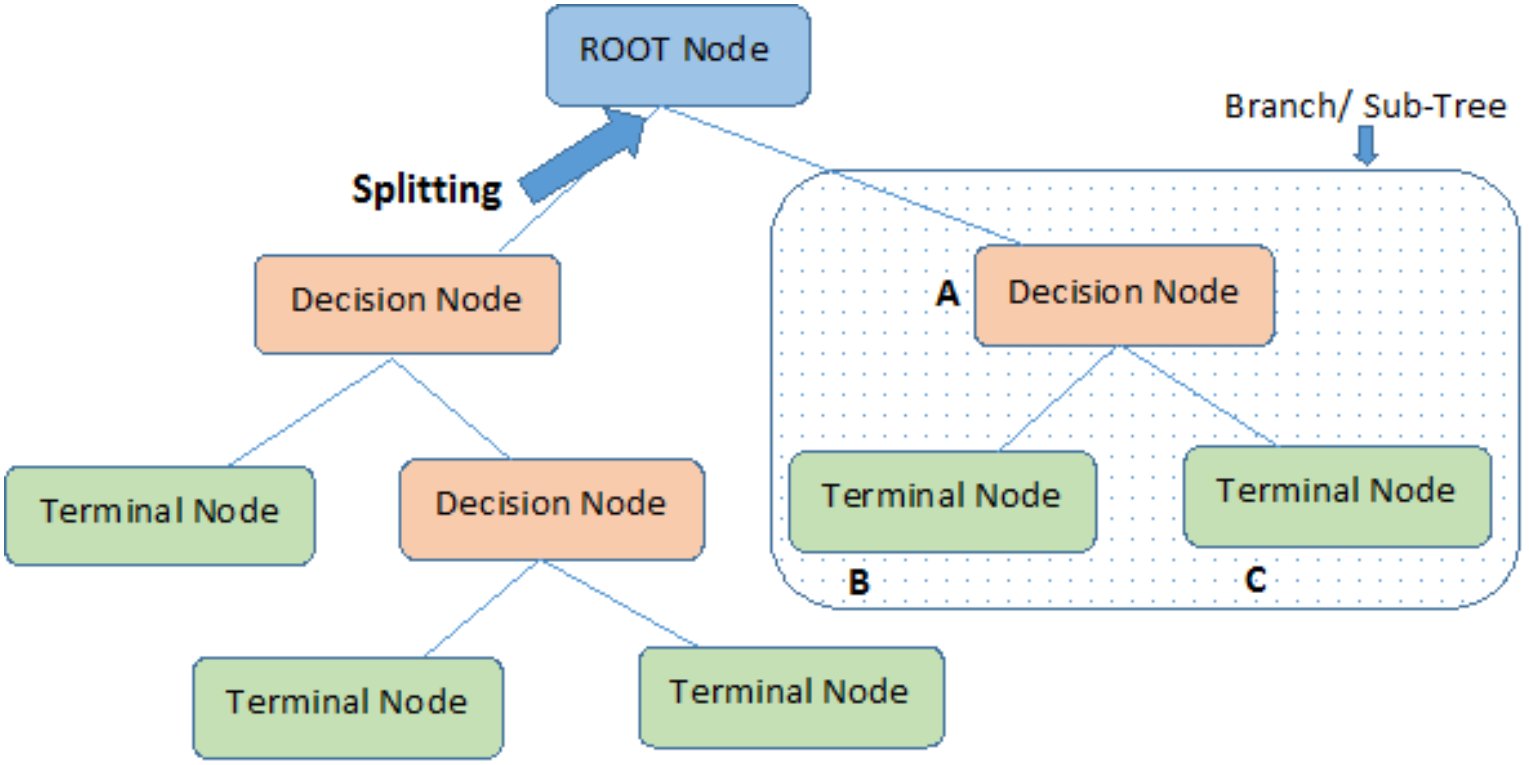
A Simple Regression Tree





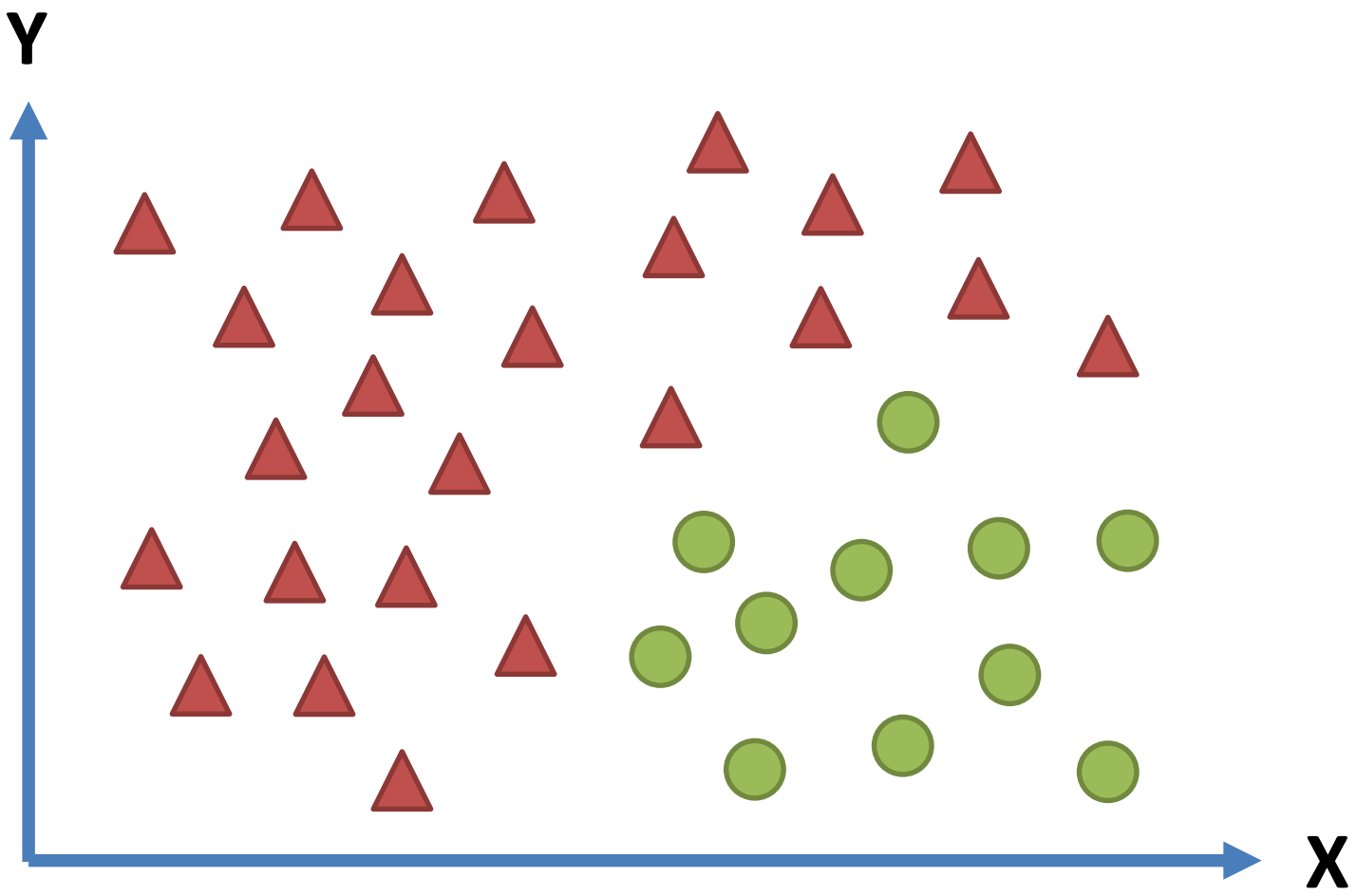
https://www.datacamp.com/community/tutorials/decision-trees-R

A General Representation of Decision Tree

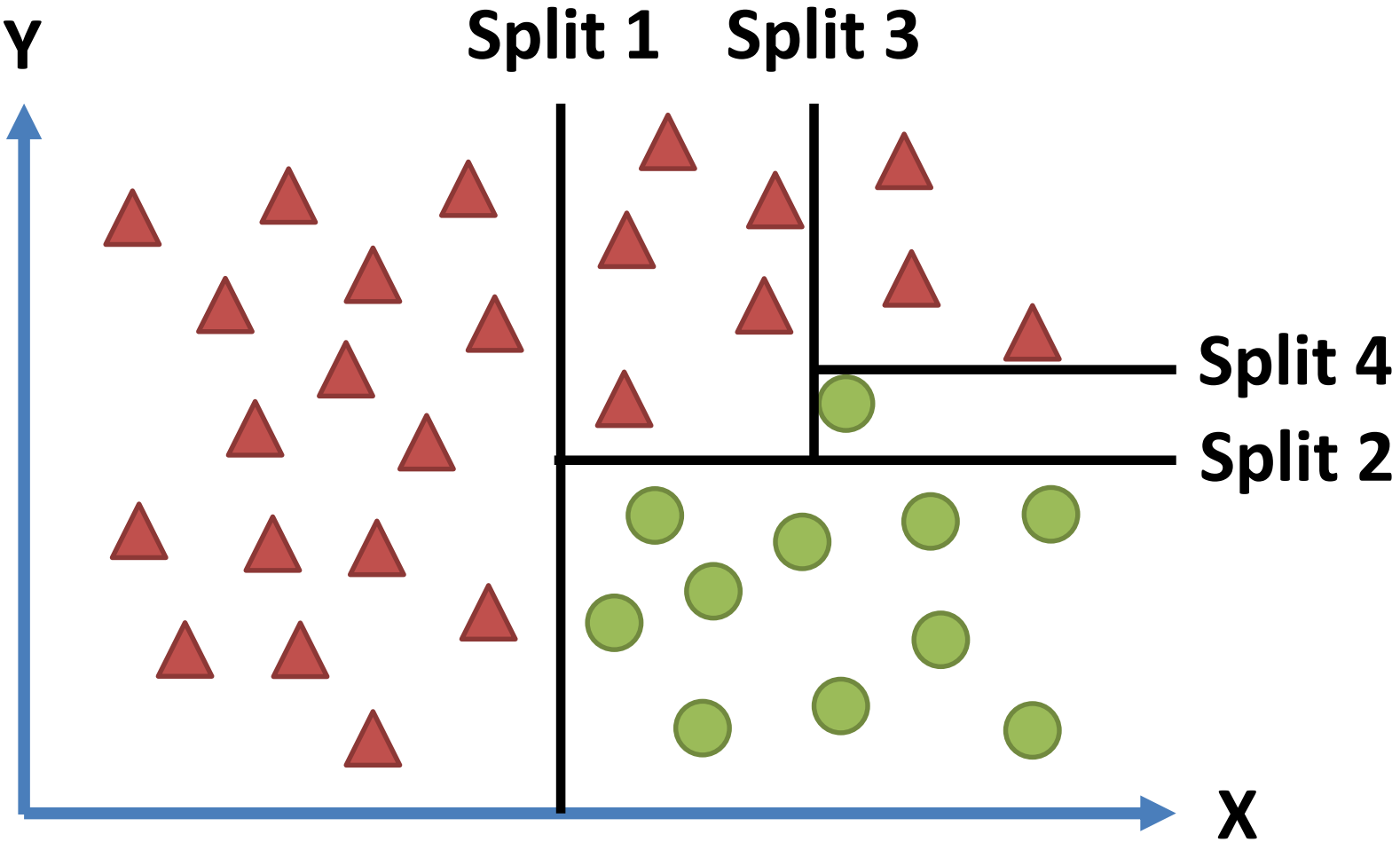


https://www.datacamp.com/community/tutorials/decision-trees-R

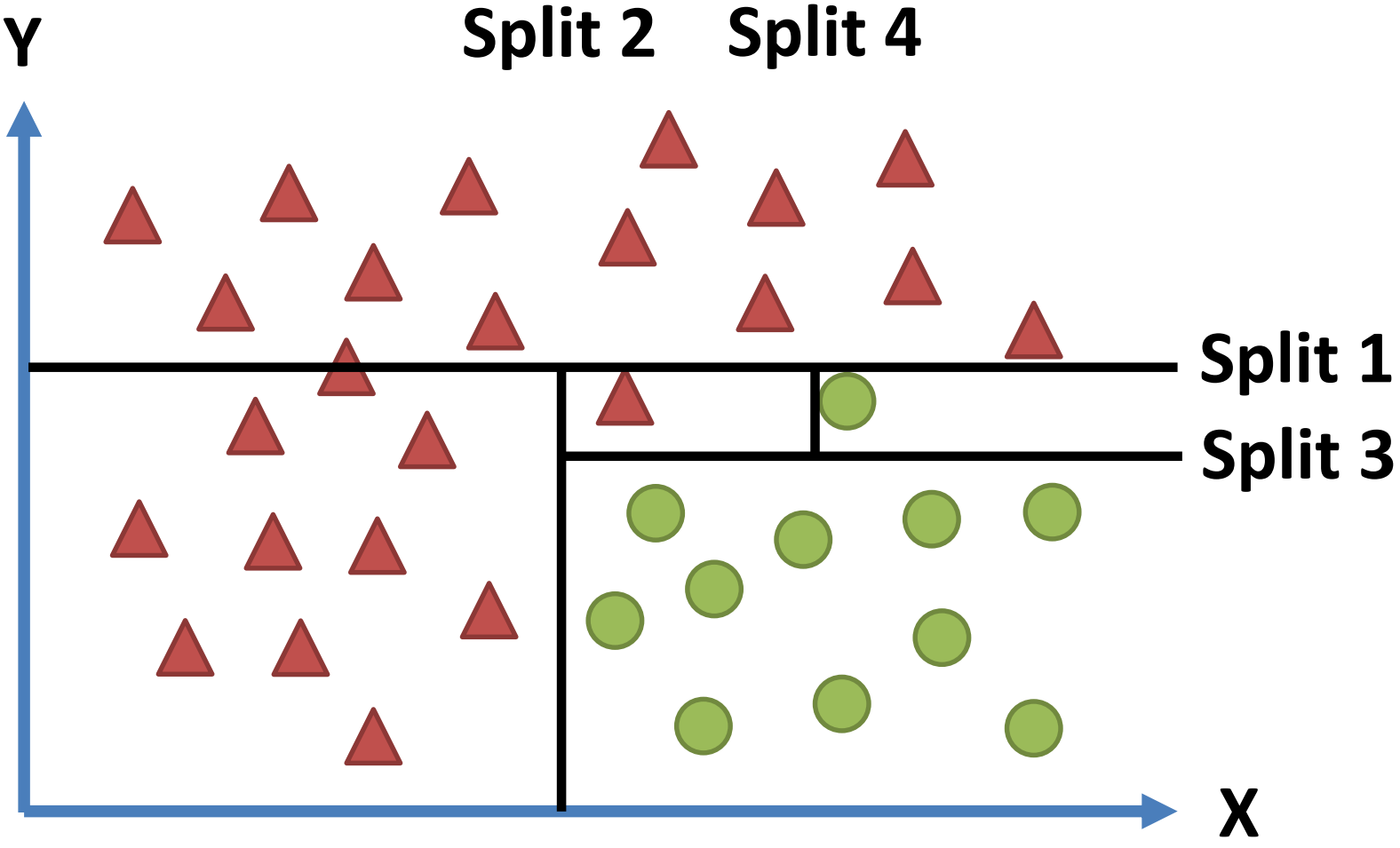
Problem: Many Ways to Fit Data



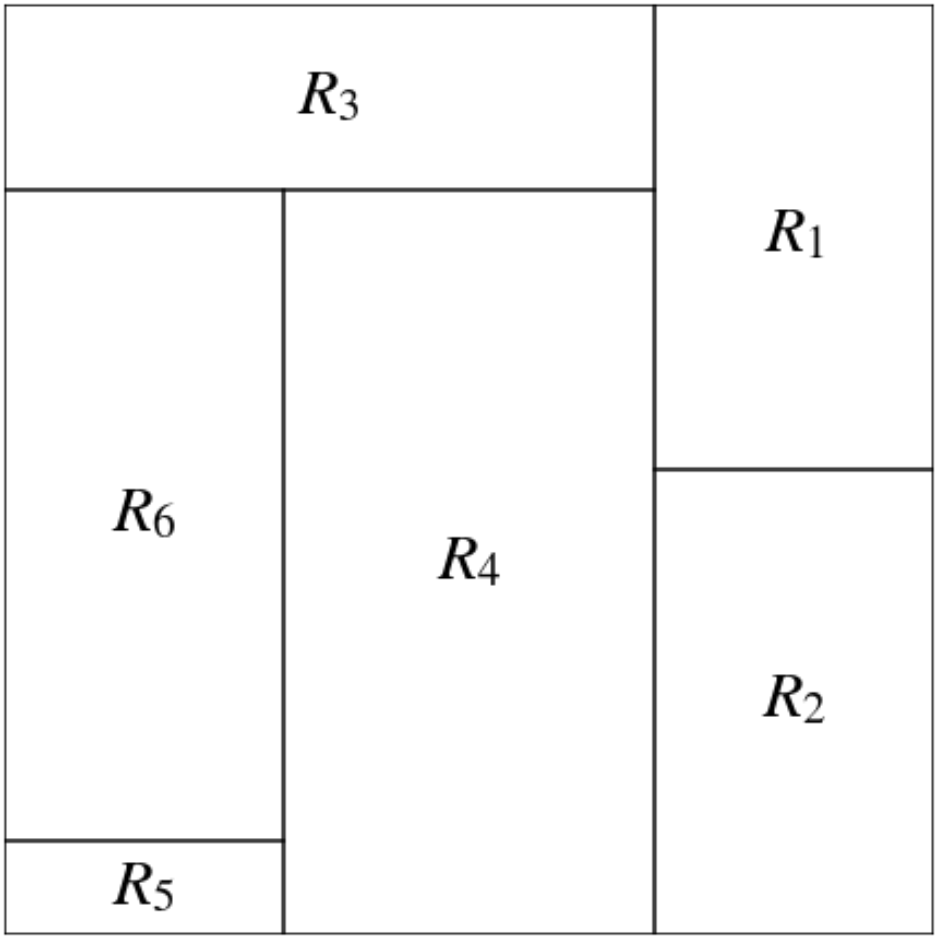
One Way to Do Classification Tree



Another Way to Do Classification Tree

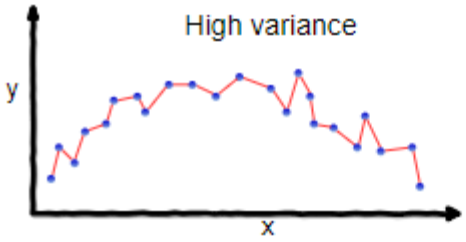


Problem: Overfitting

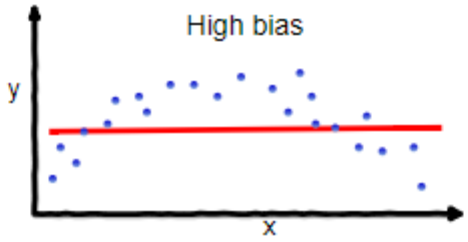


https://towardsdatascience.com/

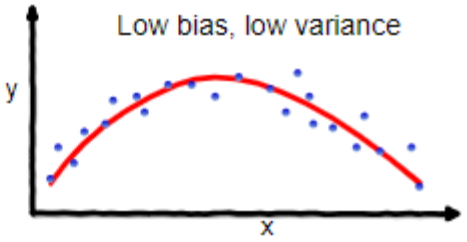
What is Overfitting?



overfitting



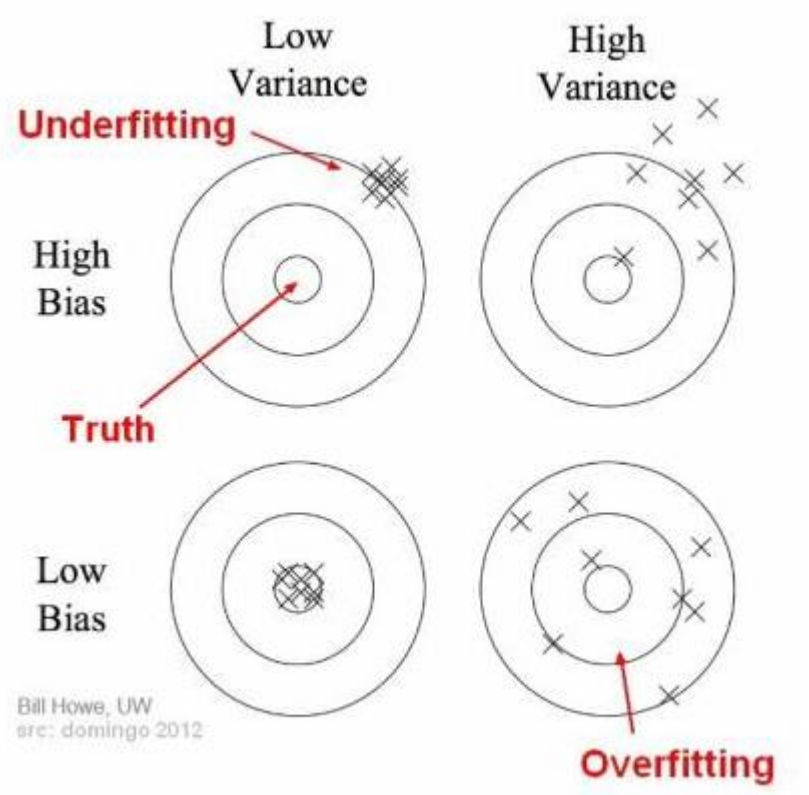
underfitting

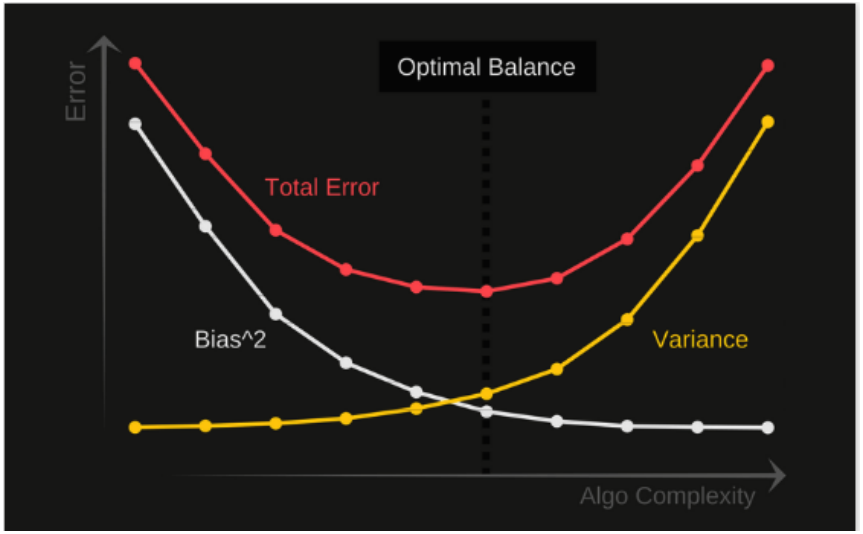


Good balance

https://towardsdatascience.com/

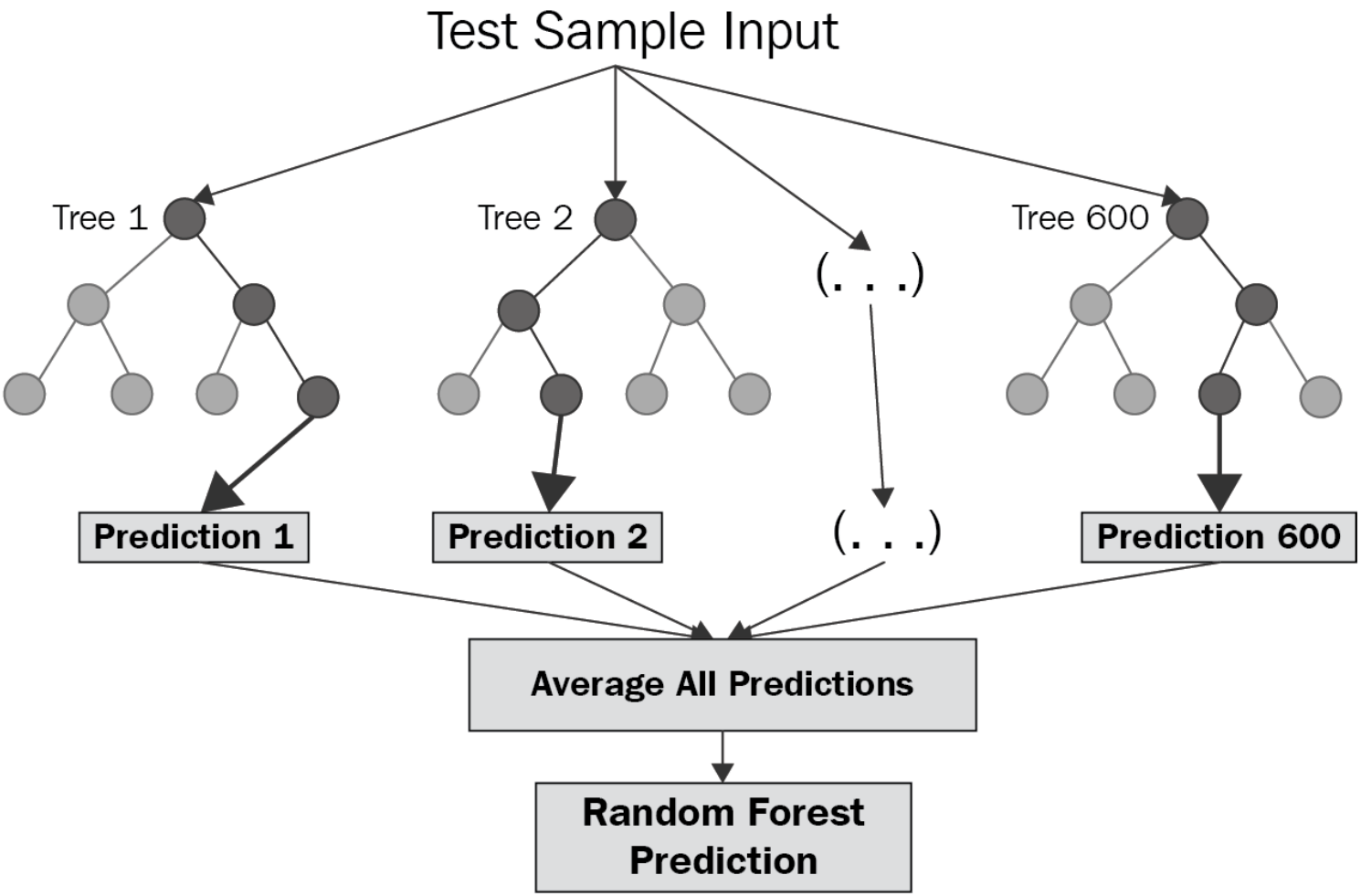
Variance vs. Bias





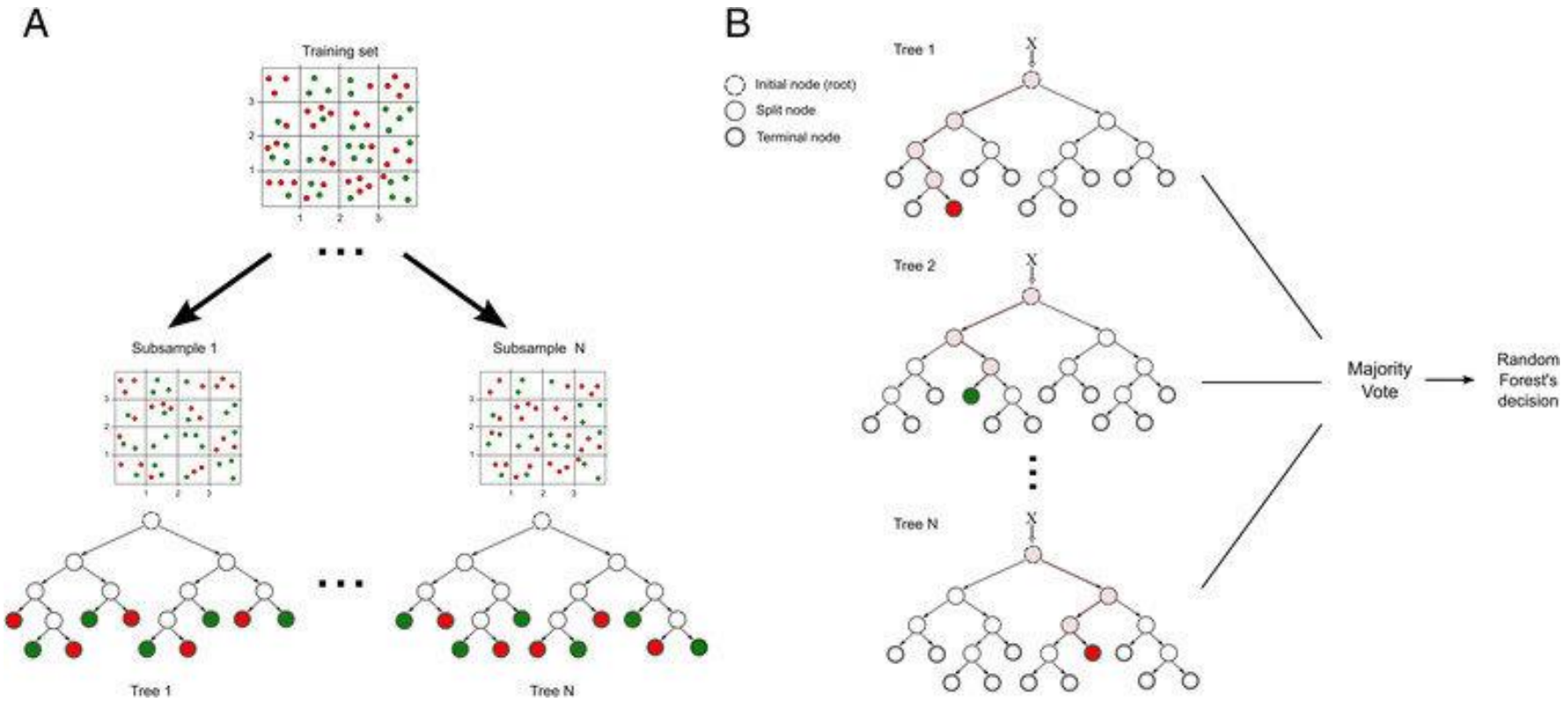
https://towardsdatascience.com/

Solution…Ensemble Methods



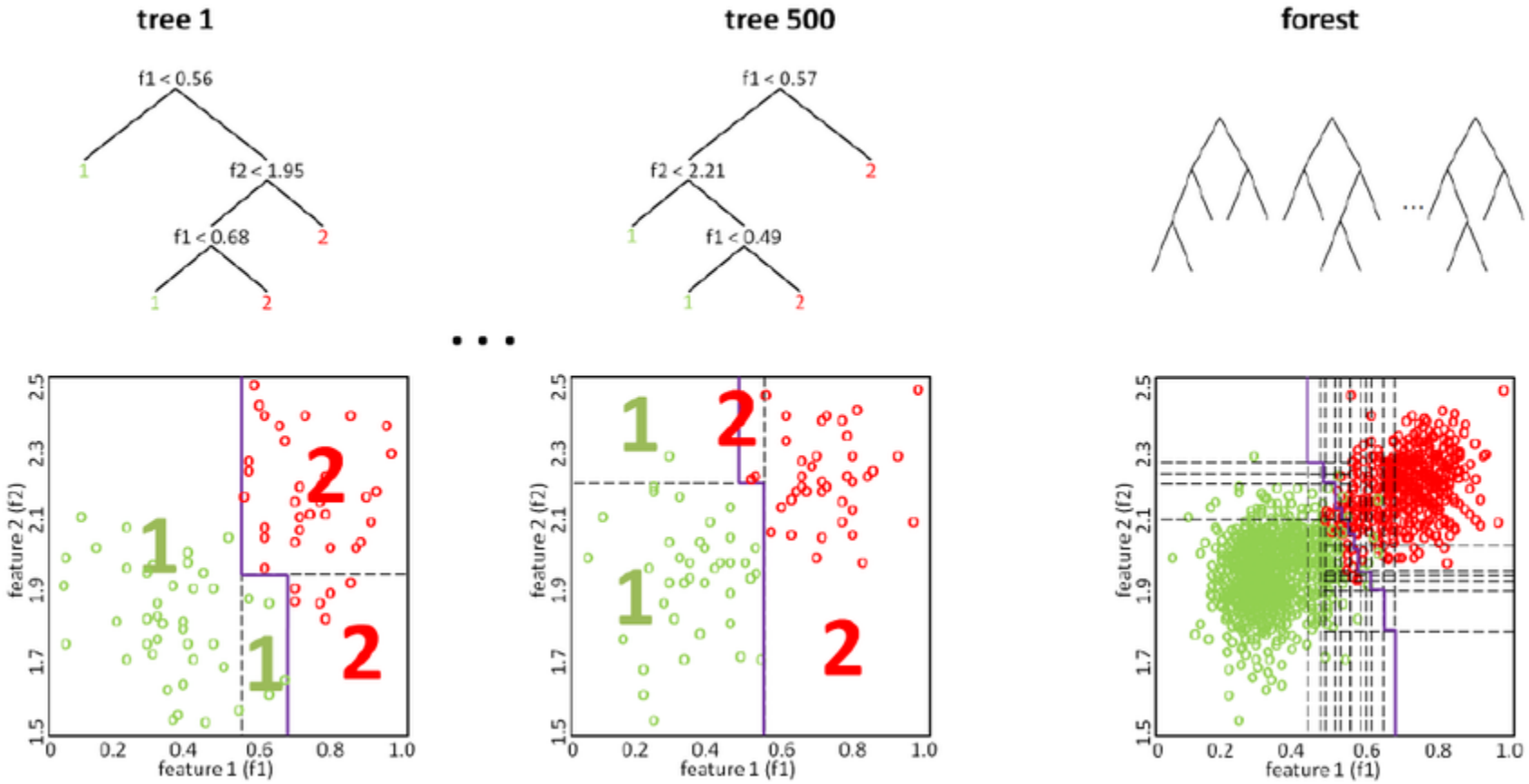
https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781789132212/2/ch02lvl1sec21/decision-tree-based-ensemble-methods

Random Forest Decision Tree



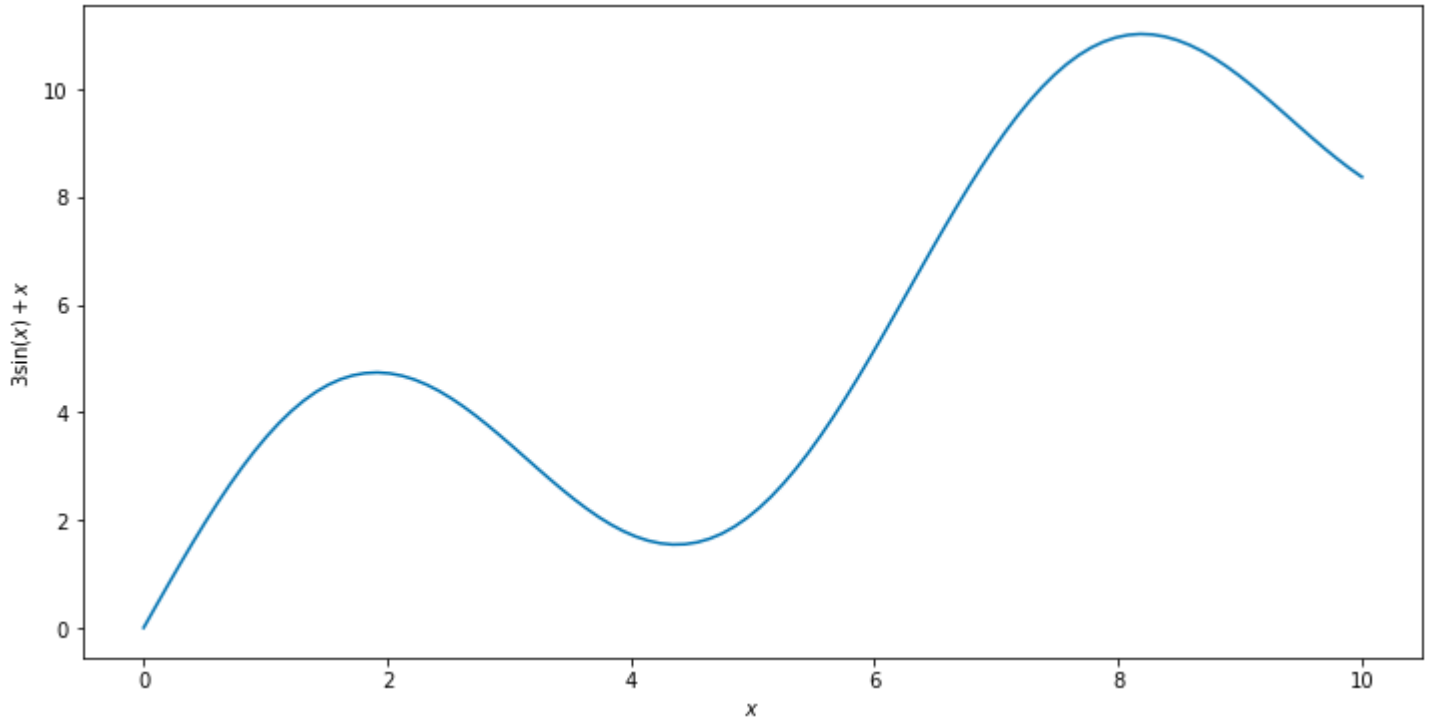
https://www.researchgate.net/figure/Random-forest-model-Example-of-training-and-classification-processes-using-random_fig5_280533599

(Random Forest) Decision Tree



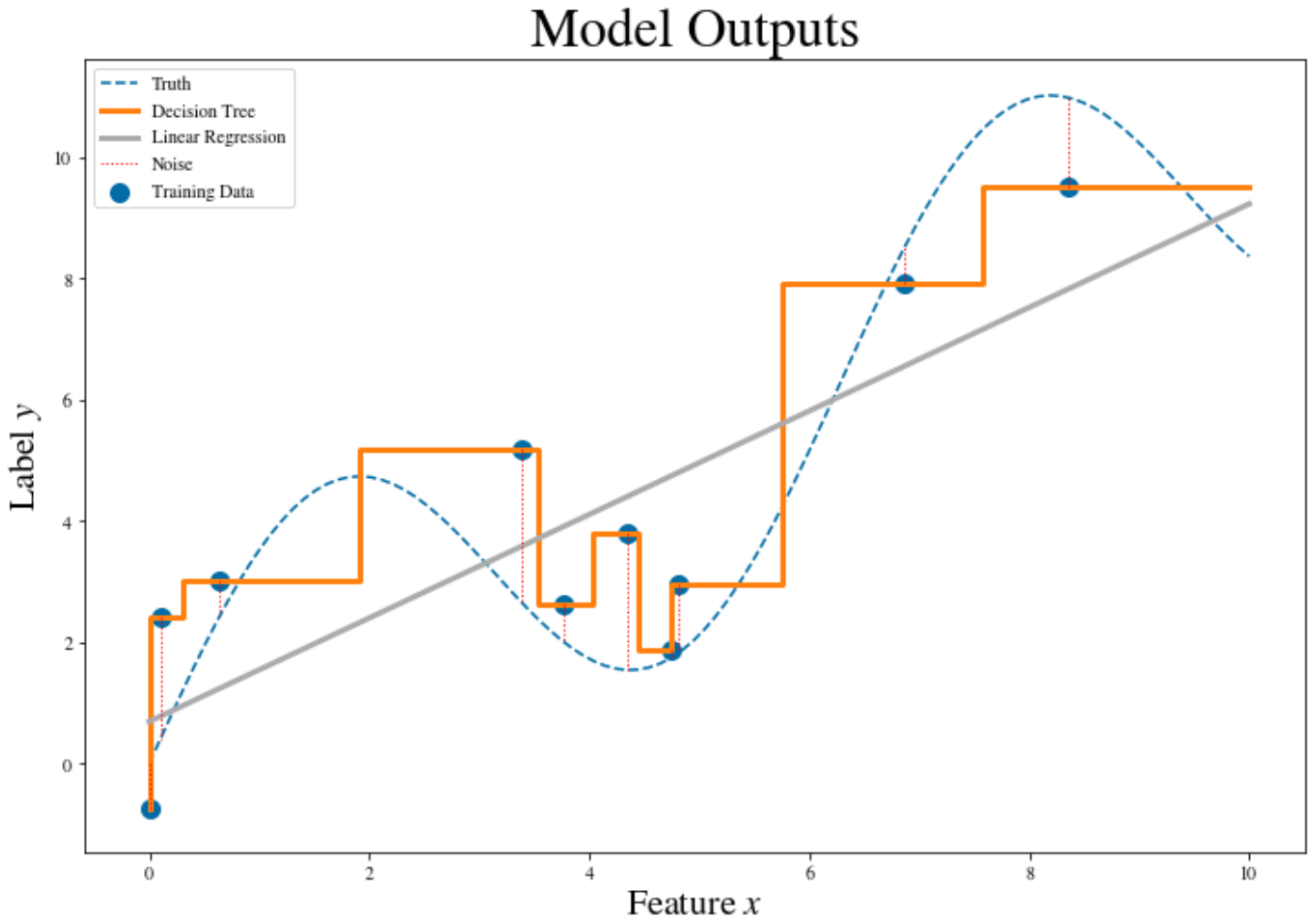
A Simple Example

$$y = h(x) = 3\sin(x) + x + \varepsilon \text{ with } \varepsilon \sim \mathcal{N}(0, 1)$$



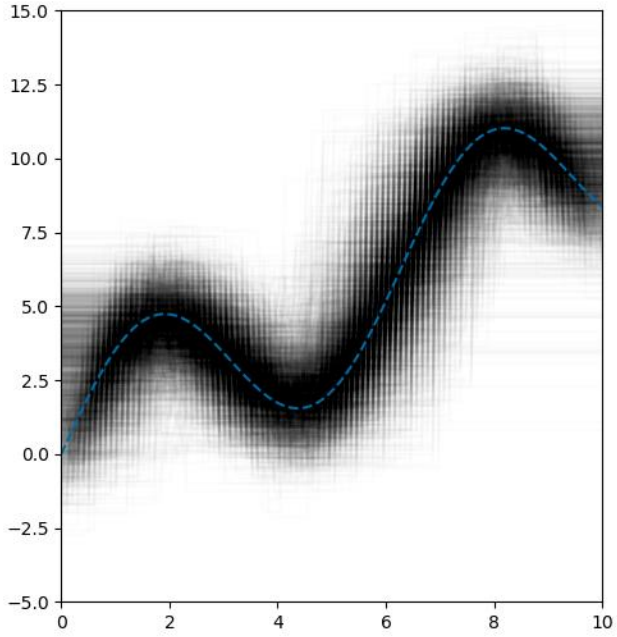
https://towardsdatascience.com/

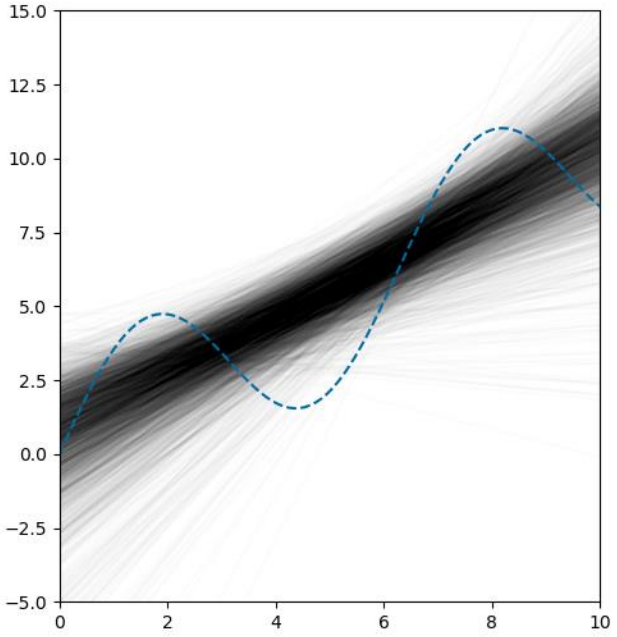
Decision Tree vs. Linear Regression



https://towardsdatascience.com/

Decision Tree vs. Linear Regression

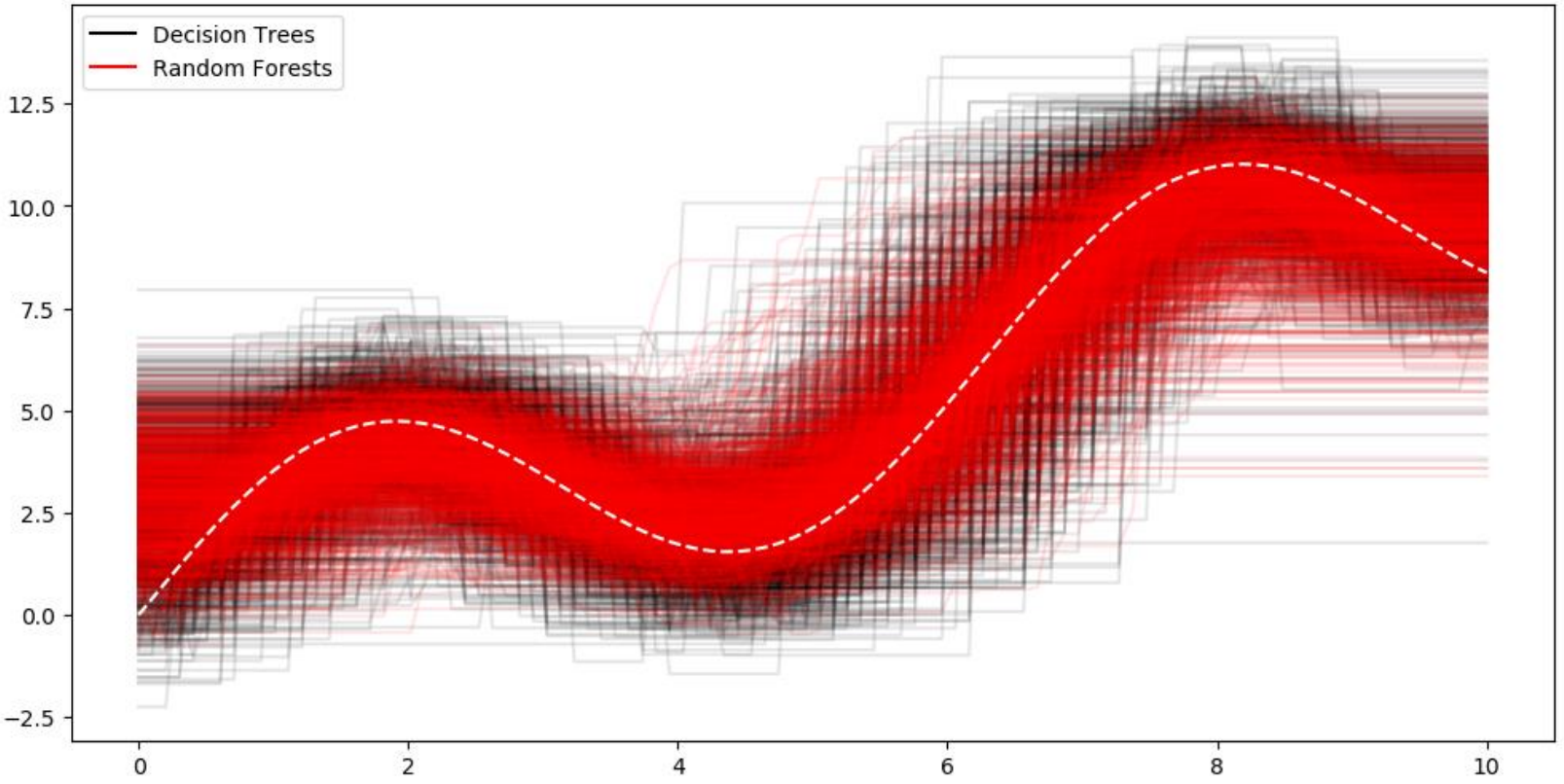




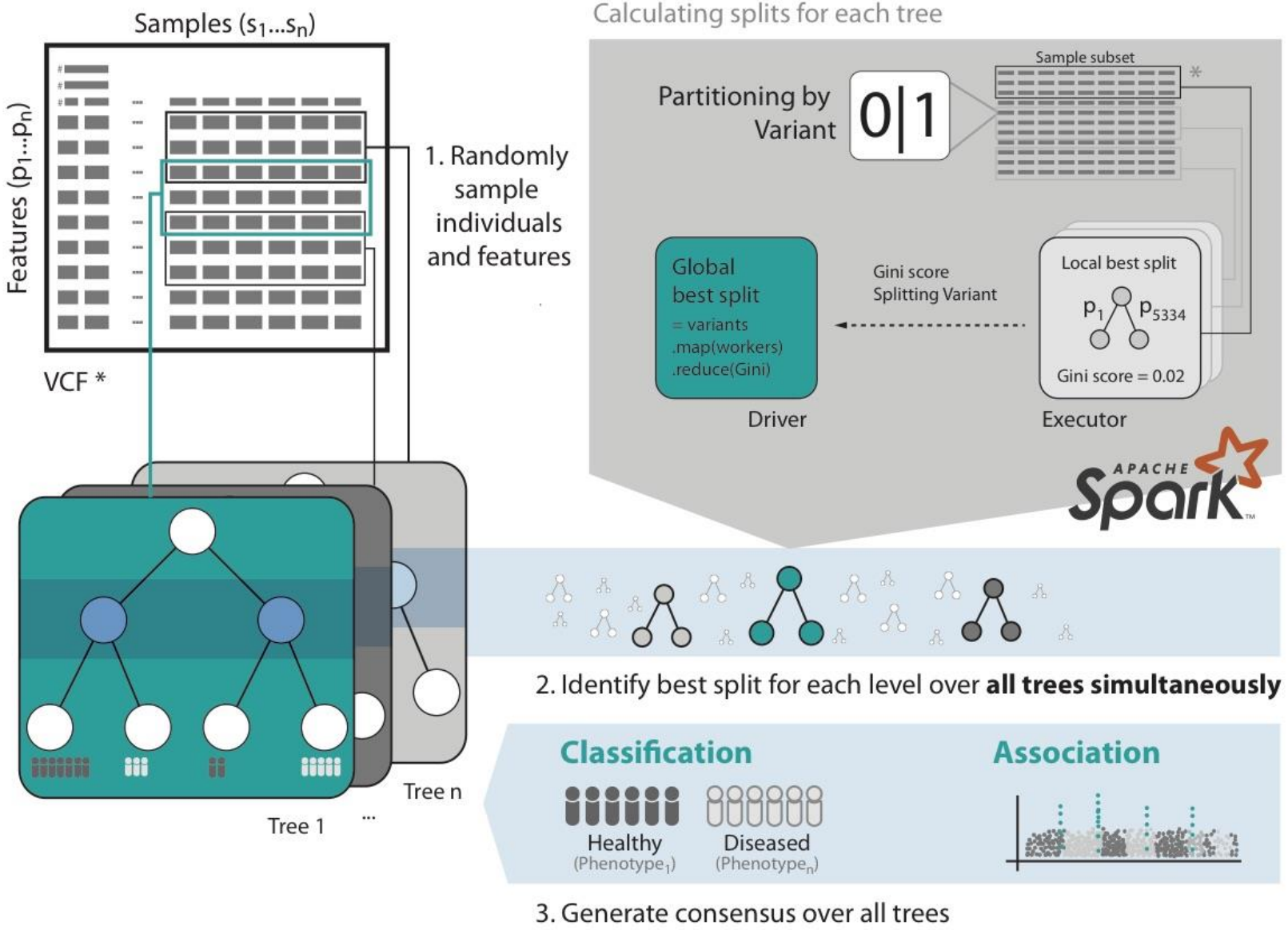
Algorithm	Bias	Variance
Decision Tree	Low	High
Linear Regression	High	Low

https://towardsdatascience.com/

Random Forest Decision Tree



https://towardsdatascience.com/



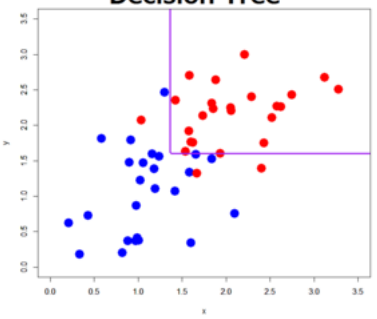
* note the transposed annotation of VCF files compared to standard ML annotation

https://databricks.com/

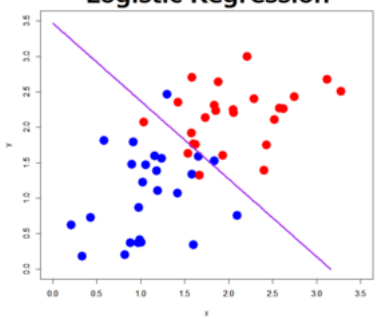
Supervised Learning Techniques

Comparisons

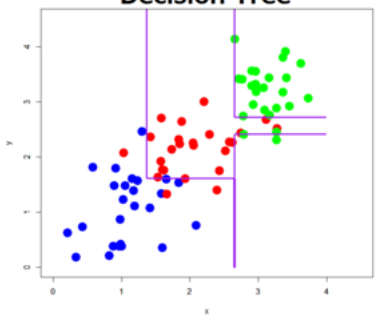
Decision Tree



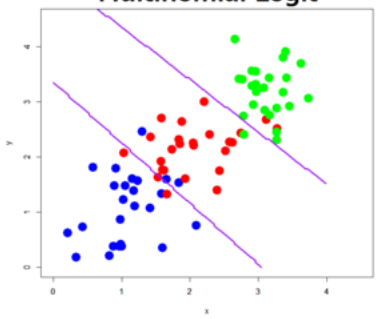
Logistic Regression



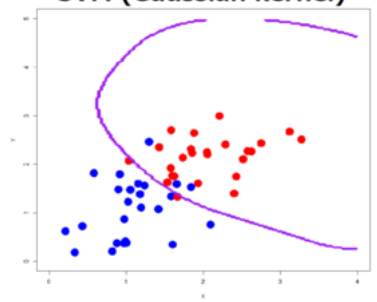
Decision Tree



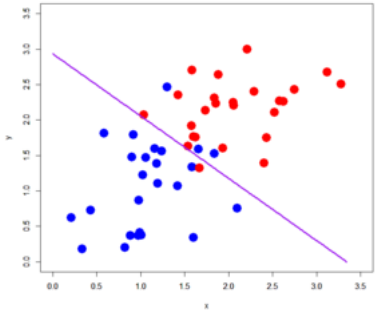
Multinomial Logit



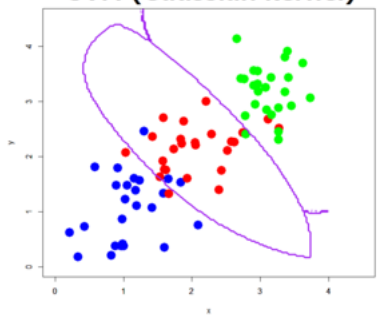
SVM (Gaussian kernel)



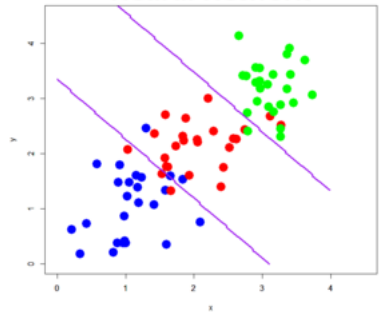
Neural Network



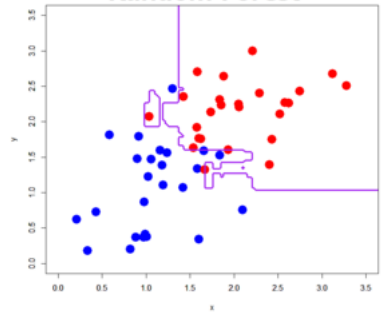
SVM (Gaussian kernel)



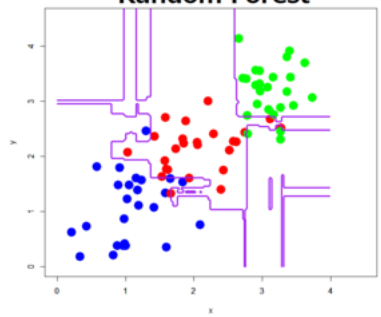
Neural Network



Random Forest



Random Forest



Summary of Decision Tree

- Advantage:
  - Easy to explain.
  - Can be displayed graphically.
  - Can easily handle qualitative predictors.

- Limitation:
  - Relatively low level of predictive accuracy
  - A small change in the data can cause a large change in the final estimated tree.

- By aggregating many decision trees, using methods like *bagging*, *random forests*, and *boosting*, the predictive performance of decision trees can be substantially improved.