Machine Learning

Session 4 - Decision trees and ensemble methods


hadrien.salem@centralelille.fr


introduction-to-data-science

Introduction

What did we do last time?

Course outline

Machine learning course

Session 1: Regression

Session 2: Supervised classification

Session 3: Clustering

Session 4: Decision trees and ensemble methods

Session 5: Introduction to neural networks

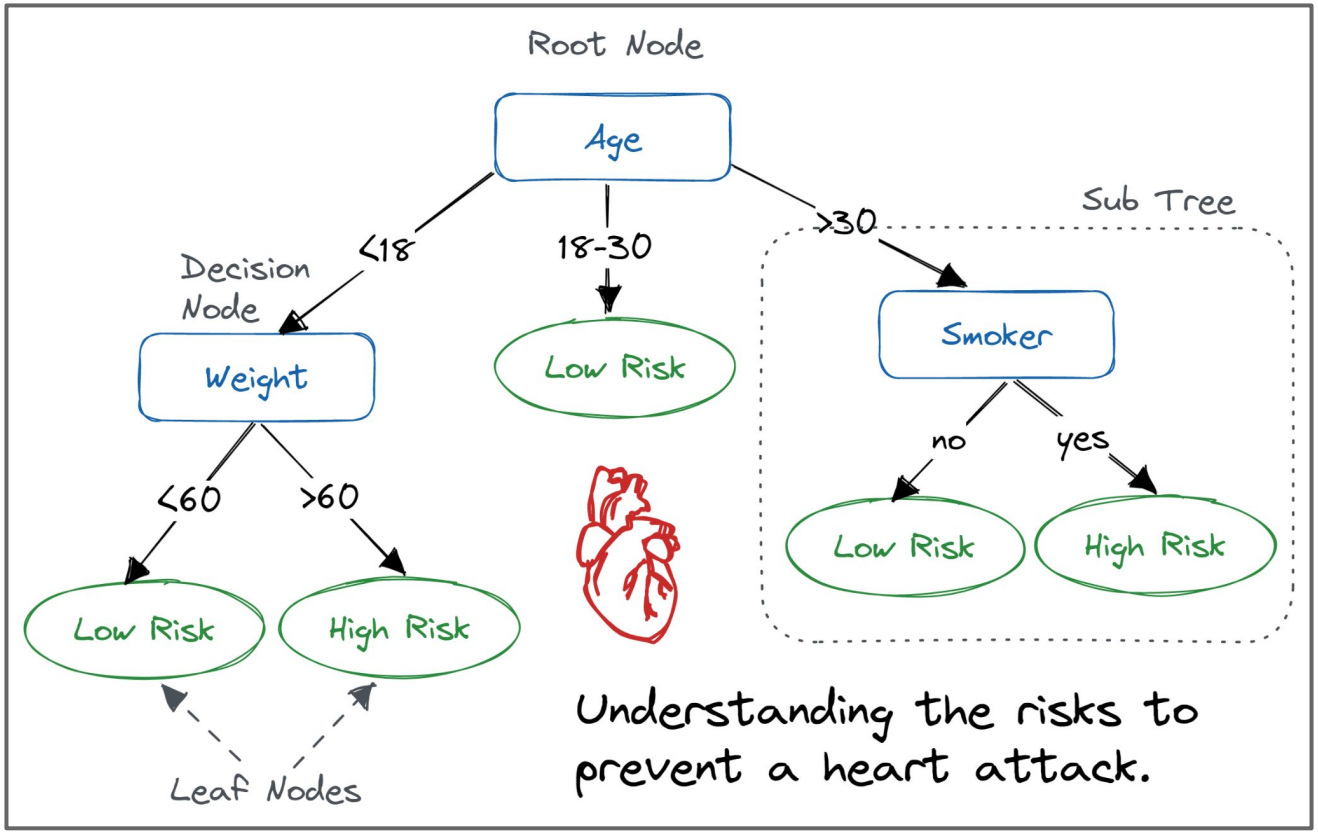
Session 6: Advanced neural networks

Session 7: Introduction to reinforcement learning

Session 8: Reading science papers

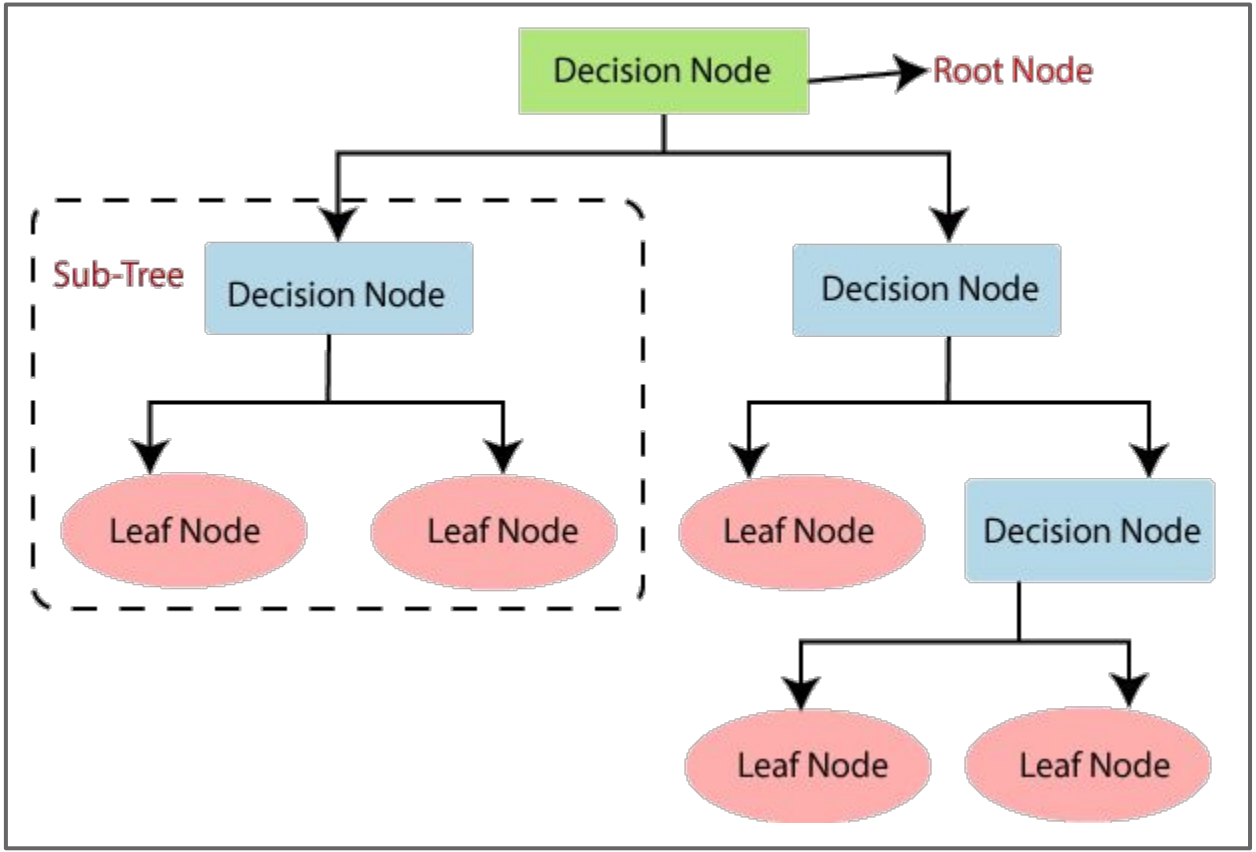
Project

What are decision trees?



Decision trees - Example

Image source



Vocabulary

Image source

Definition

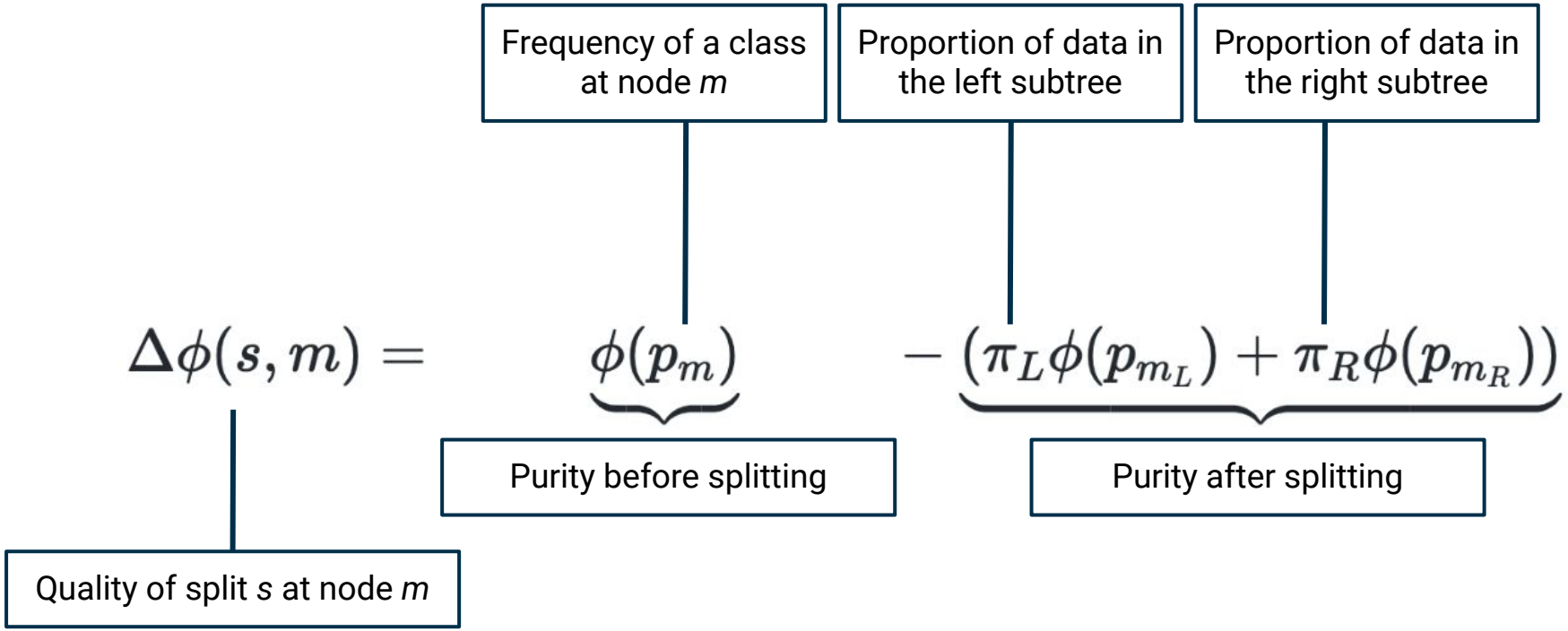
Purity of a node

A node is 100% pure when all of its data belongs to a single class.

It is 100% impure when it contains the same proportion of each class.
(e.g. 50/50 for binary classification)

Several functions can be used to compute the impurity of a node:

- **Gini Index**
- **Cross-entropy**
- **Misclassification error**



Different splits are tested recursively to find the best partitioning

Strength and weaknesses of decision trees

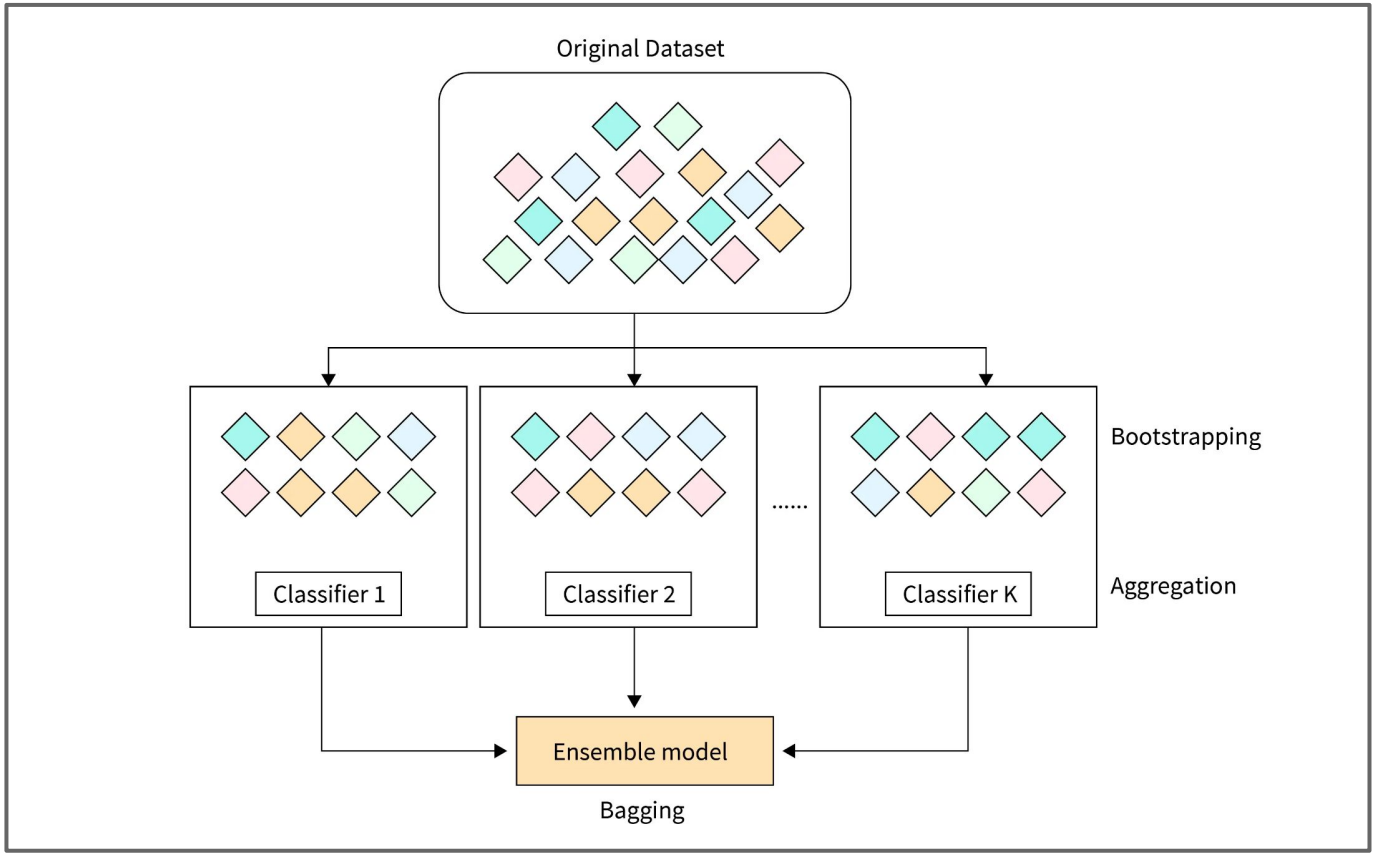
Strengths

- Flexible (few hypotheses)
- Easy to interpret (explicit rules)
- Non-linear (complex decision boundaries)

Weaknesses

- Prone to overfitting
- Unstable to noise
- Expensive on large datasets

Ensemble methods

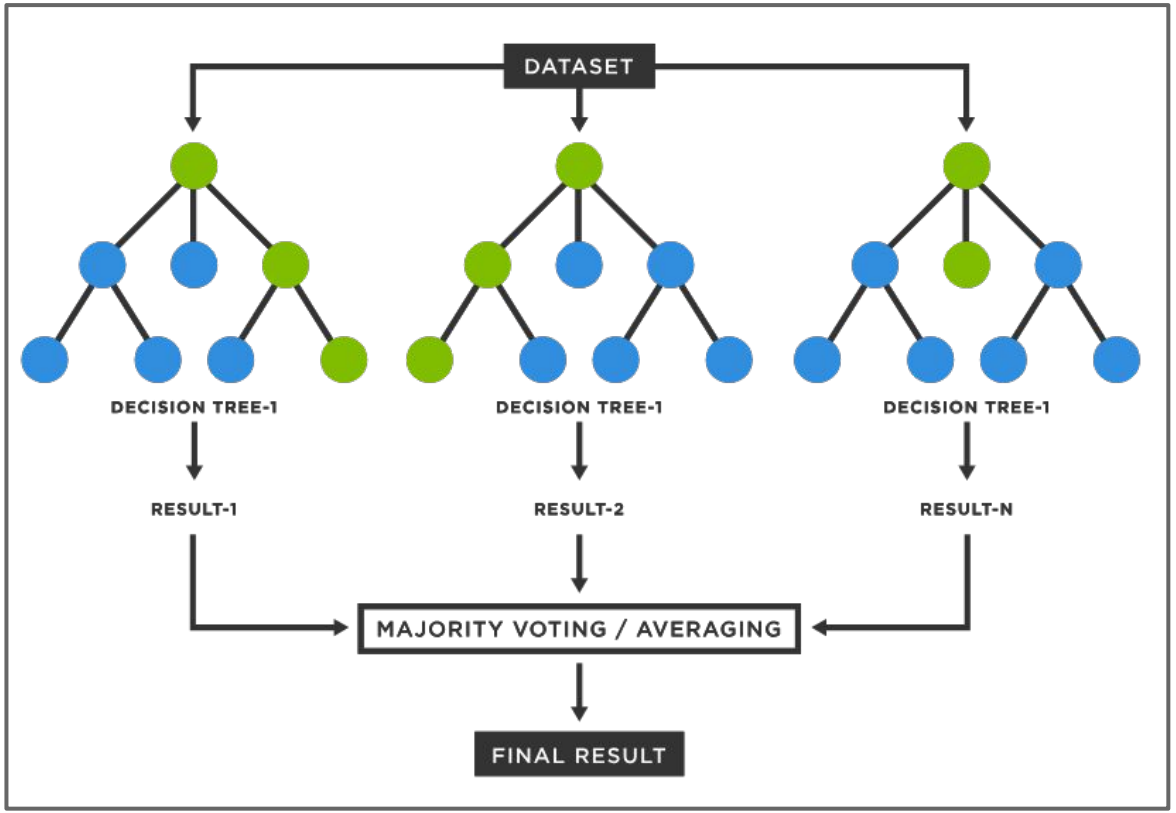


Bootstrapping
Recombining existing data to create datasets

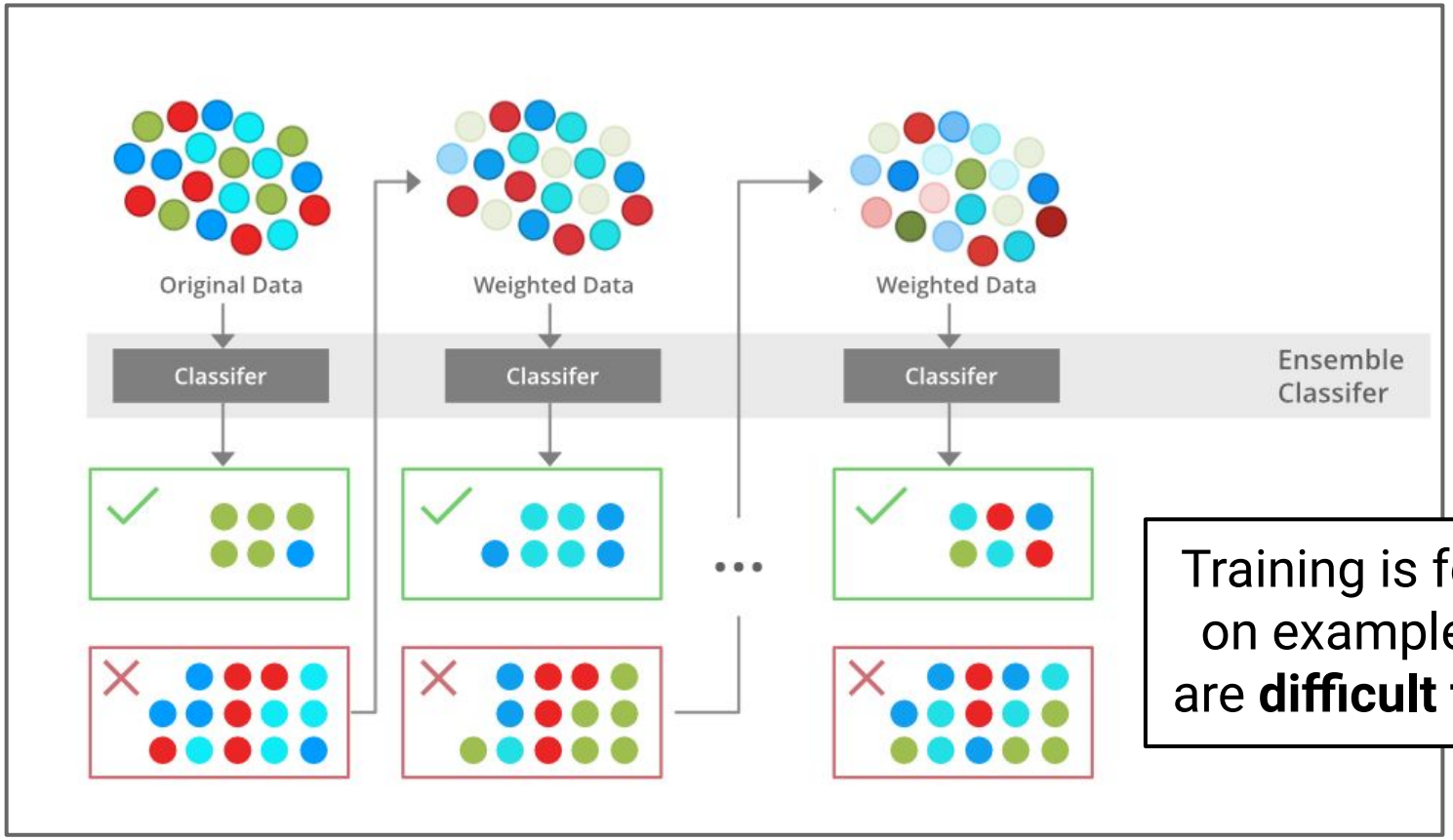
Aggregating
Training an algorithm for each dataset

Bagging (Bootstraping + Aggregating)

Image source

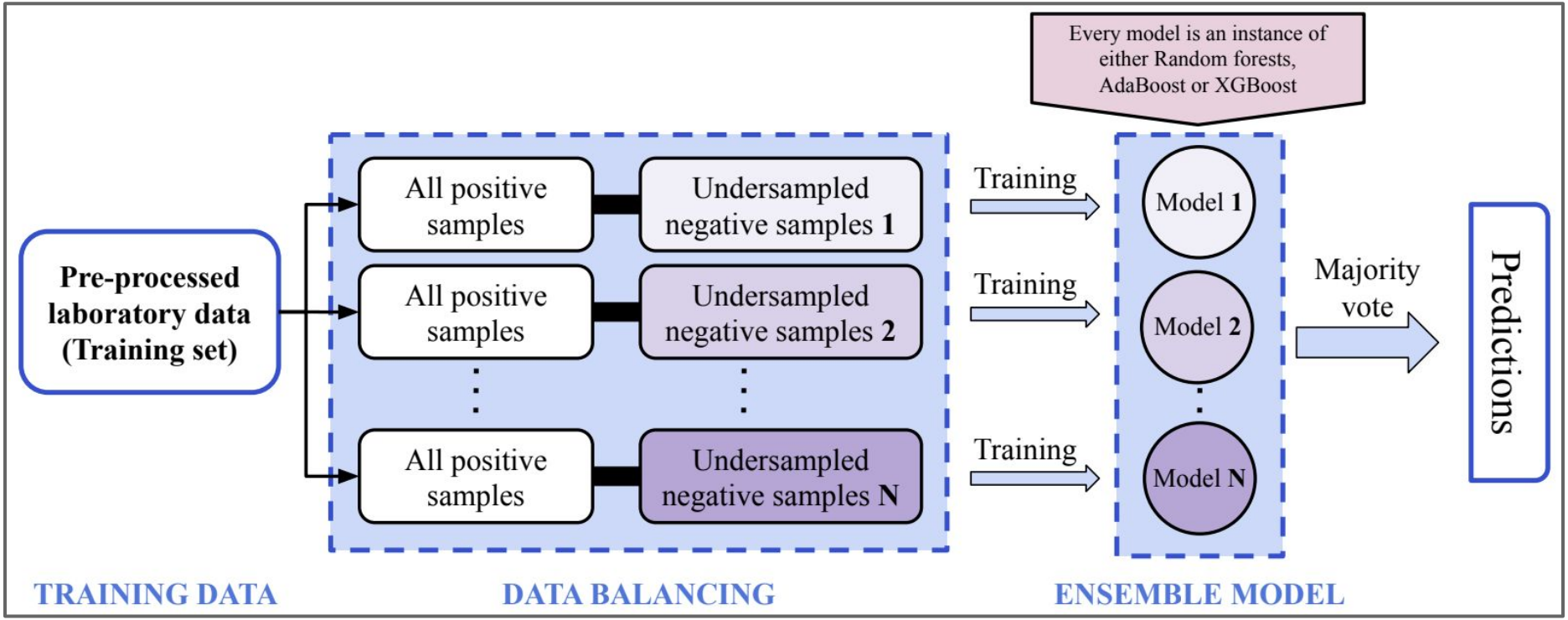


The principle is similar to bagging, except **trees are built upon random subsets of features**



Training is focused on examples that are **difficult to learn**

[Image source]



EasyEnsemble (example application to laboratory data)

Strength and weaknesses of ensemble methods

Rec

- Tends to increase accuracy
- Robust to noise
- Helps reduce overfitting

Weaknesses

- Requires more ressources
- Makes interpretation more difficult

Practical work

The notebook contains all the necessary instructions

Debrief

Debrief - MIAS

https://forms.gle/kTUwcb2yX1LTmZE29

Debrief - G3

https://forms.gle/jK7HAWRRuLHb4xBCA

Debrief

What did we learn today?

What could we have done better?

What are we doing next time?