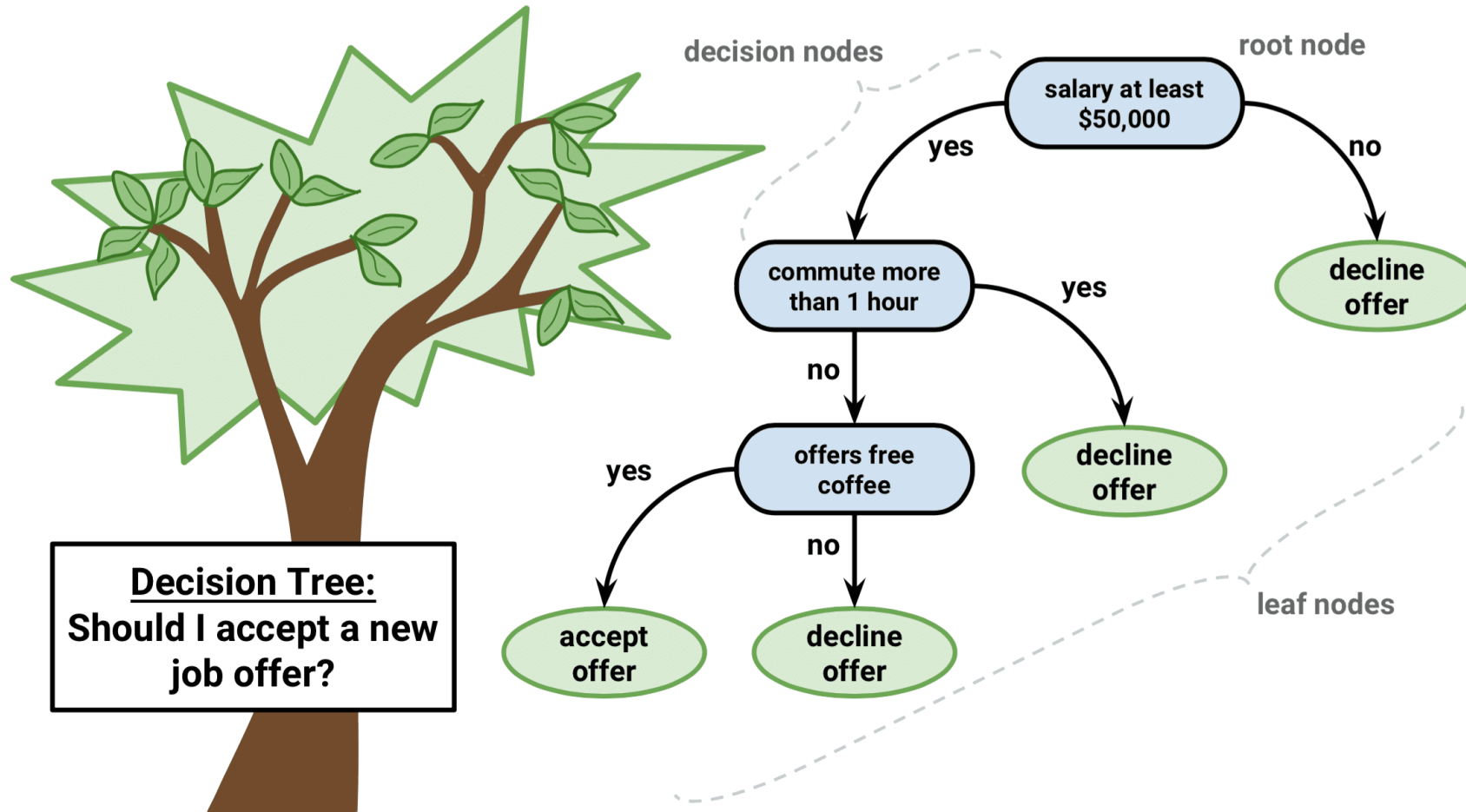# Ensemble Model: Random forest

Limitations To Decision Trees

- Decision trees tend to over fit on their training data.

Ensemble Learning

- Is a model that makes predictions based on a number of different models.

- By combining individual models, the ensemble model tends to be more flexible

- Two most popular ensemble methods are:

Bagging: Training a bunch of individual models in a parallel way.

Boosting: Training a bunch of individual models in a sequential way. Each individual model learns from mistakes made by the previous model.

Ensemble Model: Random forest

Ensemble Model
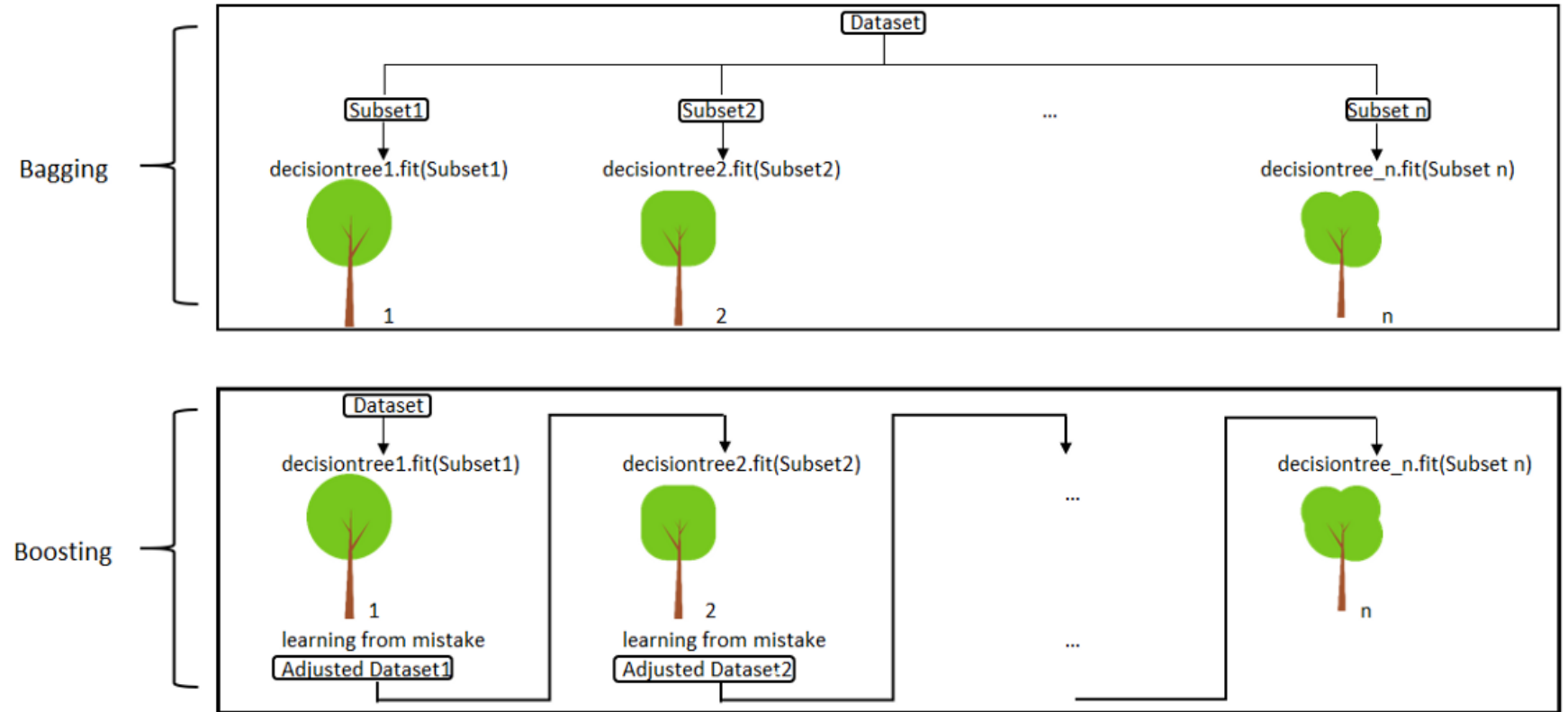
Ensemble Model: Random forest
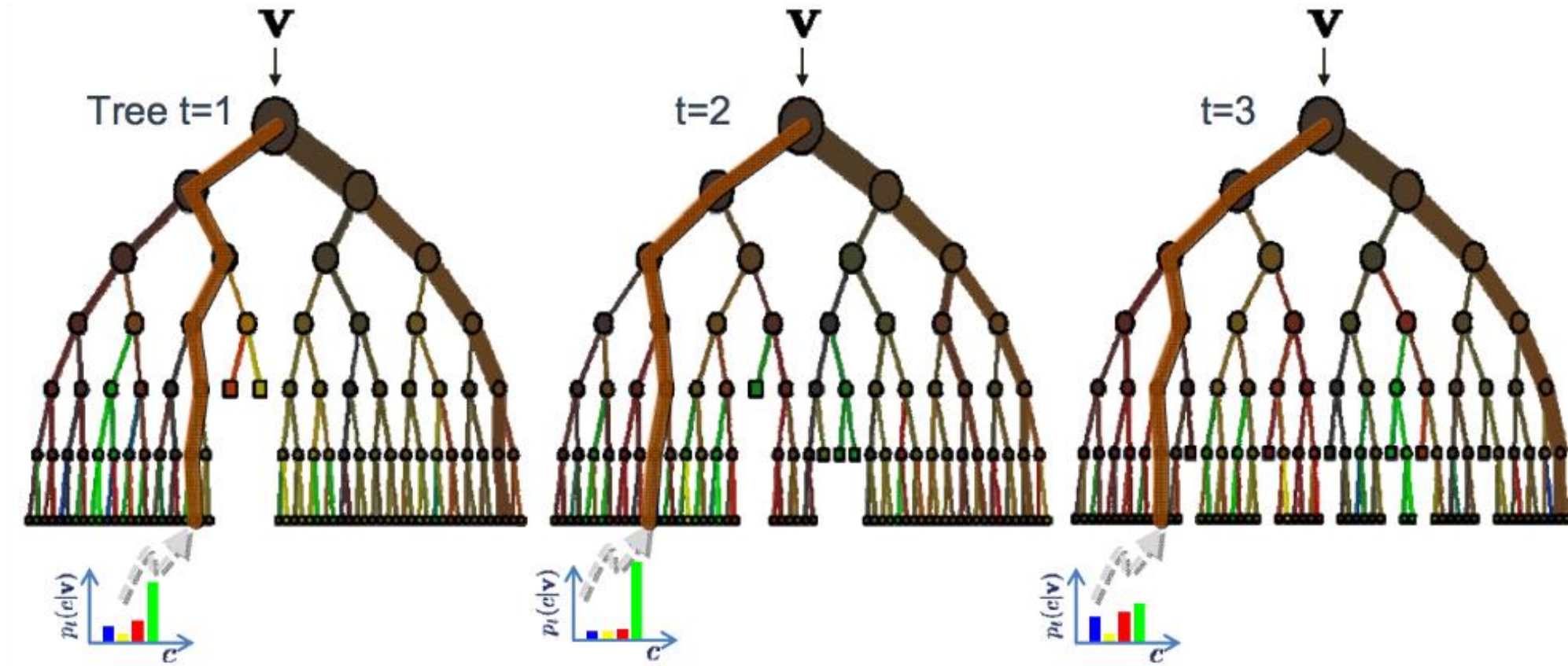
How it works

- Random Forest is a supervised learning algorithm.

- It creates a forest and makes it somehow random.

- The "forest" it builds, is an ensemble of Decision Trees, most of the time trained with the "bagging" method.

- The general idea of the bagging method is that a combination of learning models increases the overall result.

- Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

Real Life Analogy

- Imagine Andrew wants to decide, to which places he should travel during his vacation trip.

- He asks people for advice.

- First, he goes to a friend and where he traveled to in the past and if he liked it or not. Based on the answers, Andrew can decide where he should visit.

- This is a typical decision tree algorithm approach.

Ensemble Model: Random forest

Real Life Analogy

- Afterwards, Andrew starts asking more and more of his friends to advise him.

- Then he chooses the places that were recommend the most to him

- It is typical Random Forest algorithm approach.

Building A Random Forest

- Choose the number of trees you'd like in your forest (M)

- Choose the number of samples you'd like for each tree (n)

- Choose the number of features you'd like in each tree (f)

- For each tree in M:
  - Select n samples from all observations
  - Select f features at random
  - Train a decision tree using the data set of n samples with f features
  - Save the decision tree

Ensemble Model: Random forest

How A Random Forest Makes A Prediction

- Given an observation (o).

- For each tree (t) in the model:

  - predict the outcome (p) using the tree (t) applied to observation(o)

  - store predicted value (p) in list P

- If the model is a classifier:

  - return max_count(p)

- If the model is a regressor:

  - return avg(p)
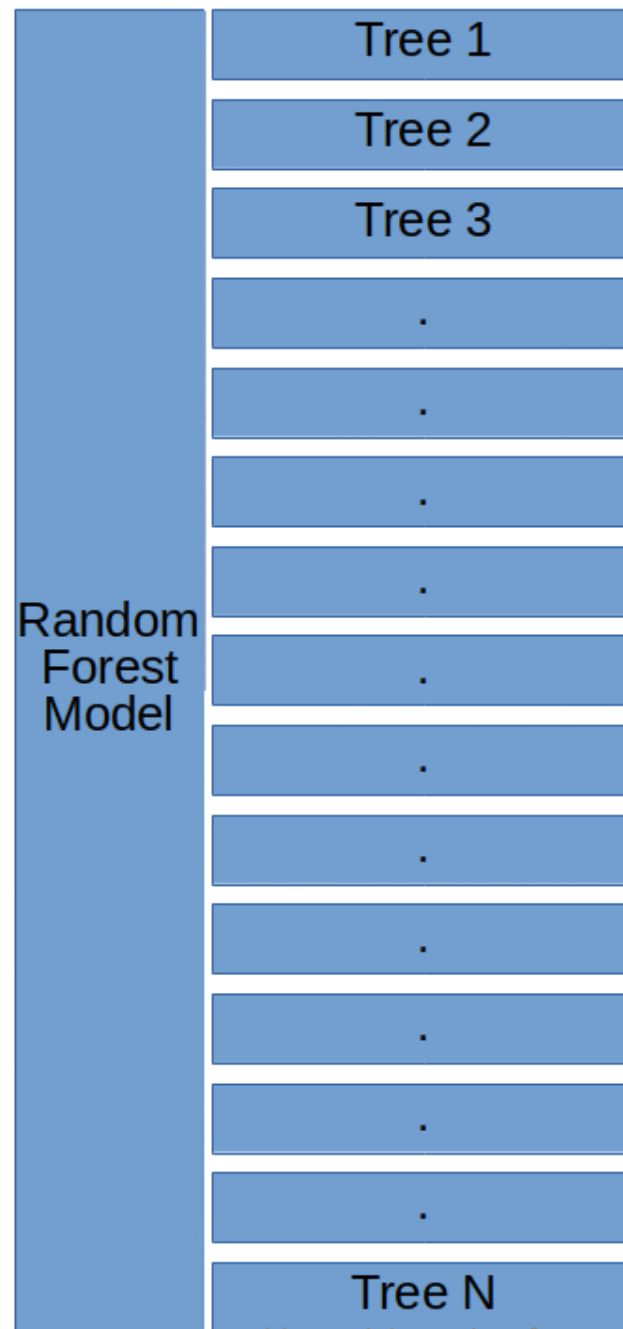
Ensemble Model: Random forest

Thanks