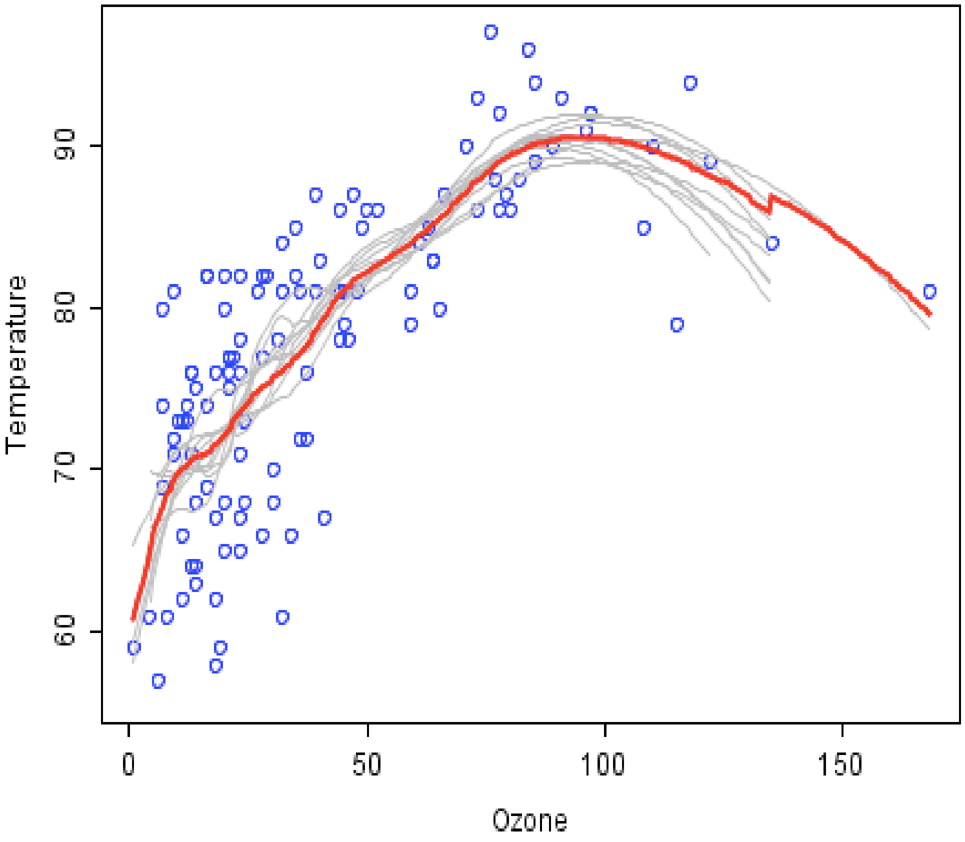
RANDOM FOREST

RANDOM Forest

- Kind of like KNN
- “Weak learners” (singles) come together to form the “Strong learner”

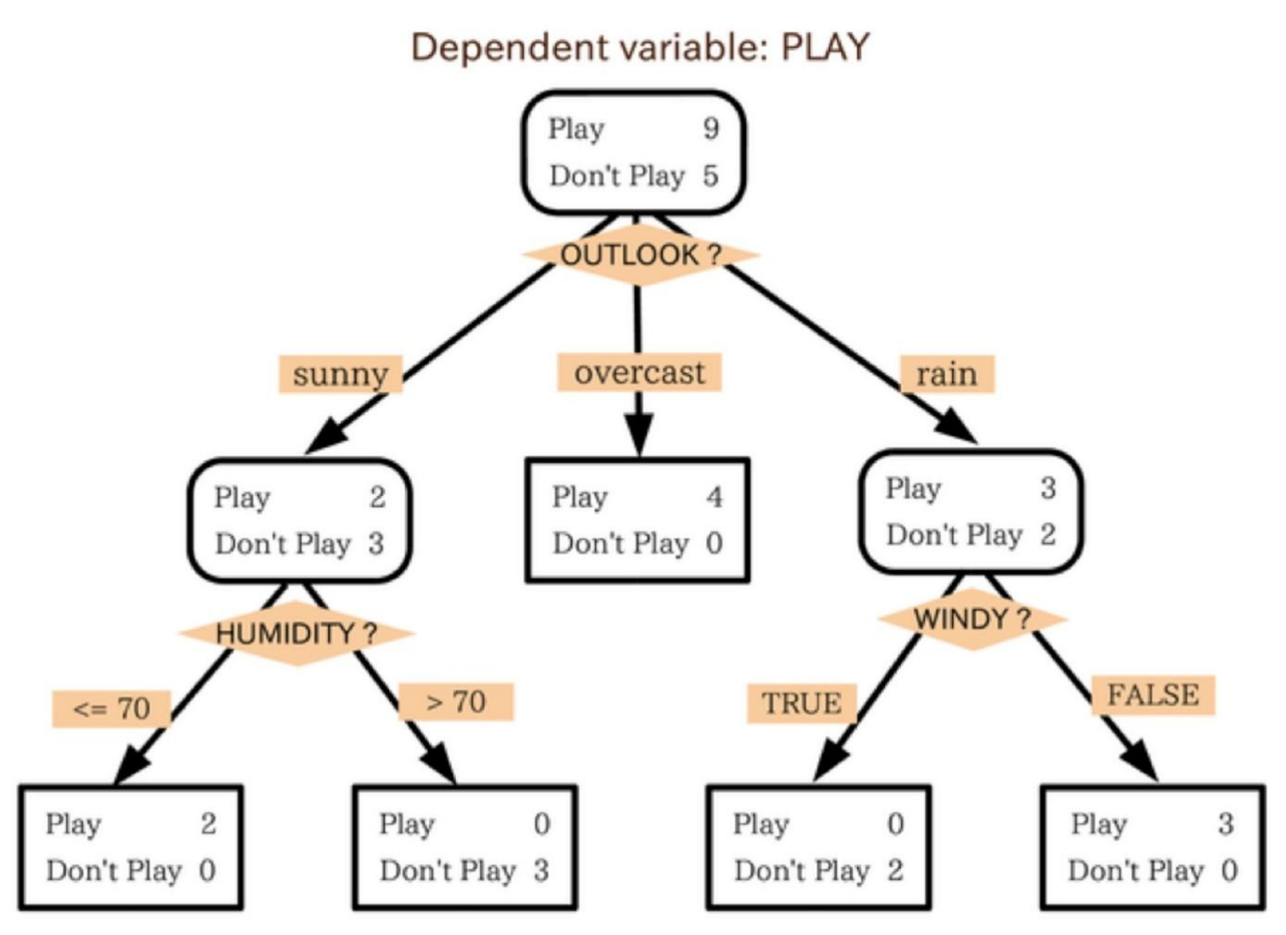
RANDOM Forest

- Blue Circles → Data
- Gray Curves → Weak Learners
- Red Curve →Strong Learner



RANDOM Forest

- Algorithm: Decision Tree (@ClassicML)



RANDOM Forest

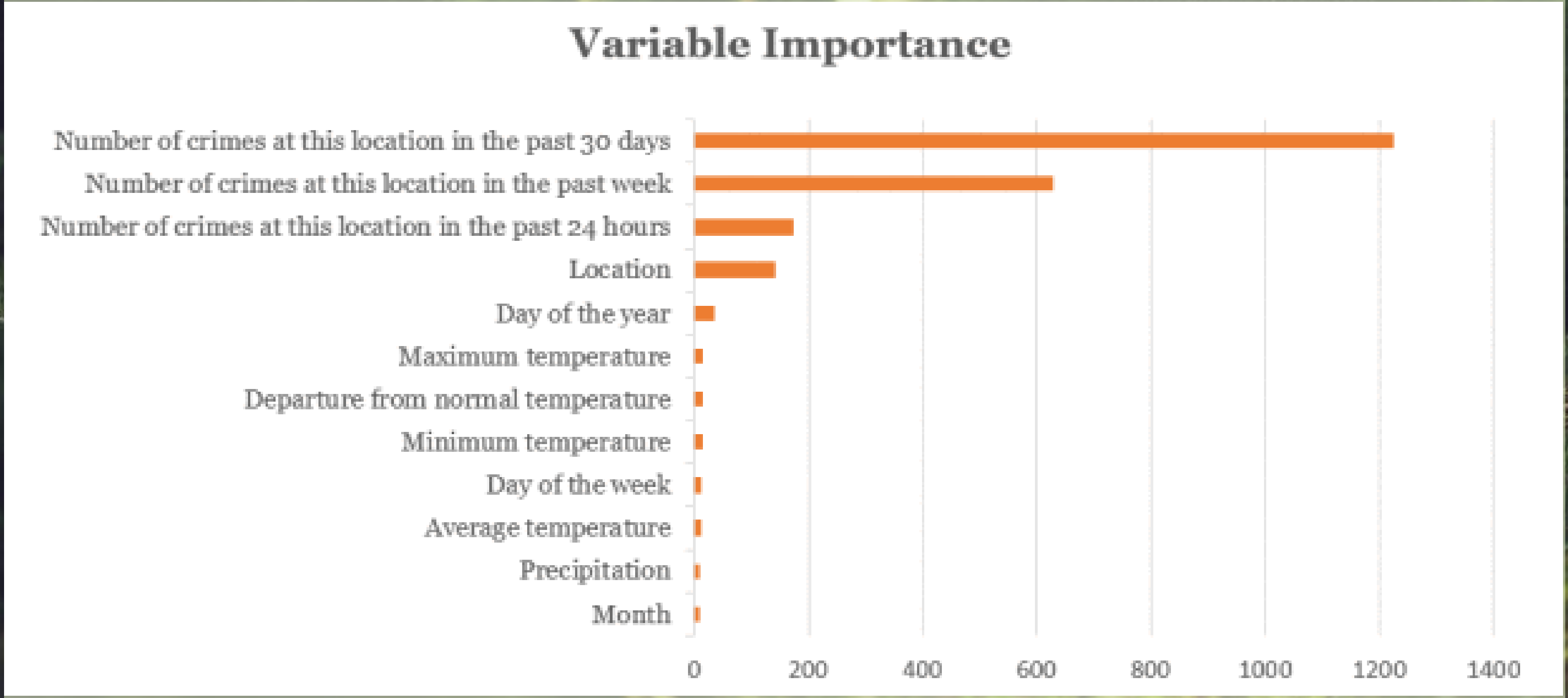
- Random FOREST combines the TREES into a FOREST
  - Tree → Weak Learner
  - Random Forest → Strong Learner

RANDOM Forest: Training

- Sample *N* cases at random with replacement to create a subset of the data.
- At each node:
  - *m* predictor variables are selected at random from all the predictor variables
  - Predictor variable with the best split makes a binary split on that node
  - Next Node: Repeat

Example: San Francisco Police Department

- Data: crime reports
- Predictor Variables:



RANDOM Forest: Strength/Weaknesses

STRENGTHS

Fast runtimes

Deal with unbalanced and missing data

WEAKNESSES

May over-fit noisy data