

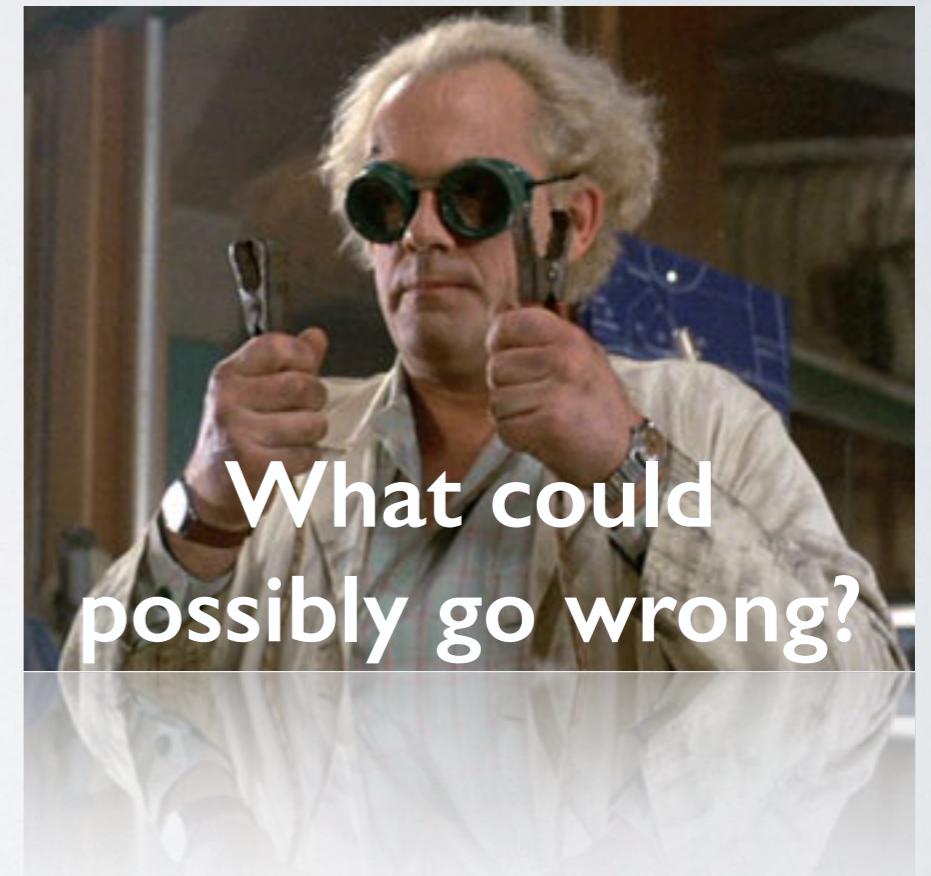
RANDOM FORESTS

Intro to Machine Learning Series



BIO

- Anthony Towry
@anthonytowry
- OSCP/CISSP/CISA
- Quant Cyber Risk Tyrannosaurus Rex
- No background in ML
 - Don't believe anything in this preso



What could possibly go wrong?

TYPES OF ML

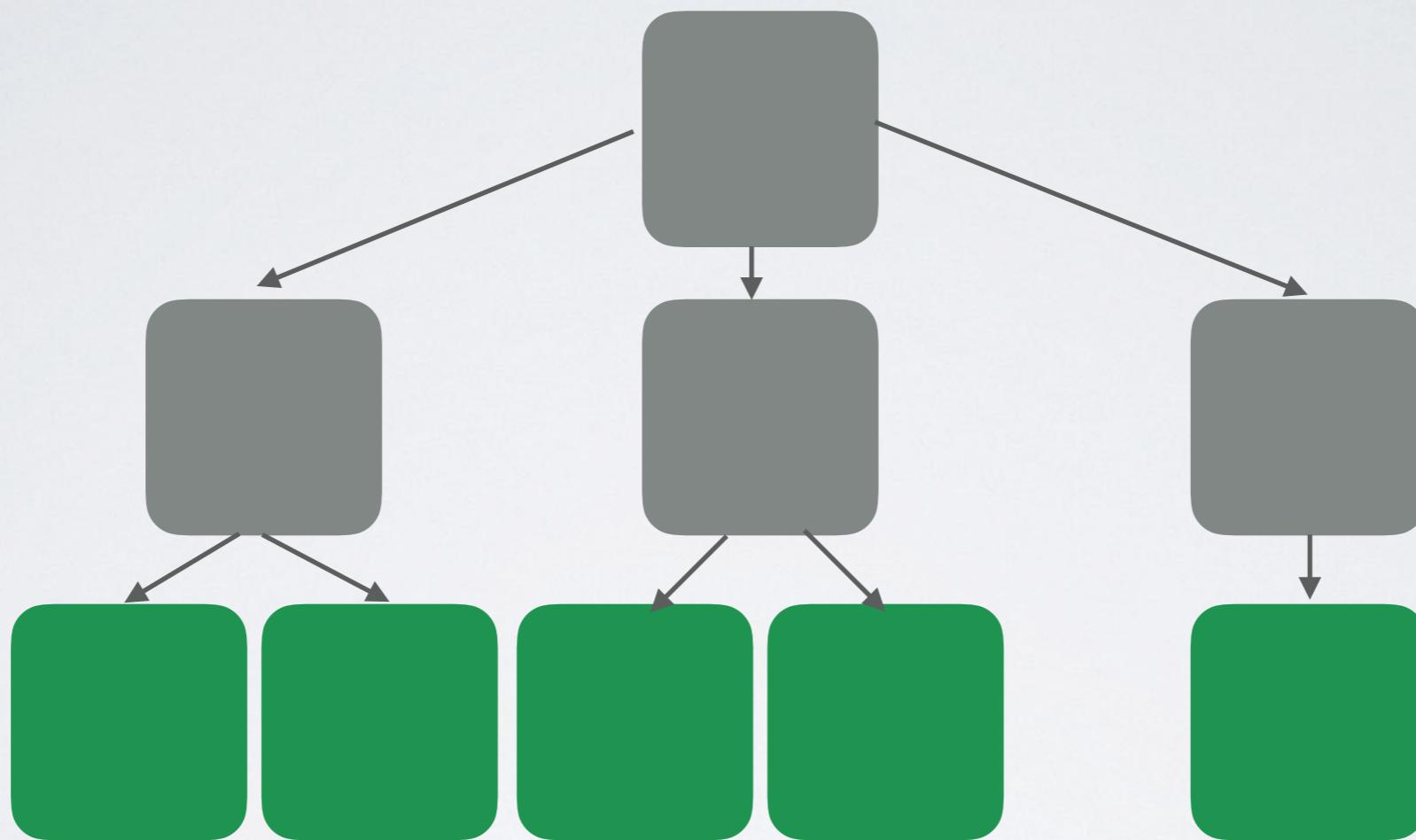
Supervised Learning

Unsupervised Learning

TREES

- Uses entropy/information gain to determine branching iteratively.
- They can handle both categorical and numeric features.

TREES



OVERFITTING

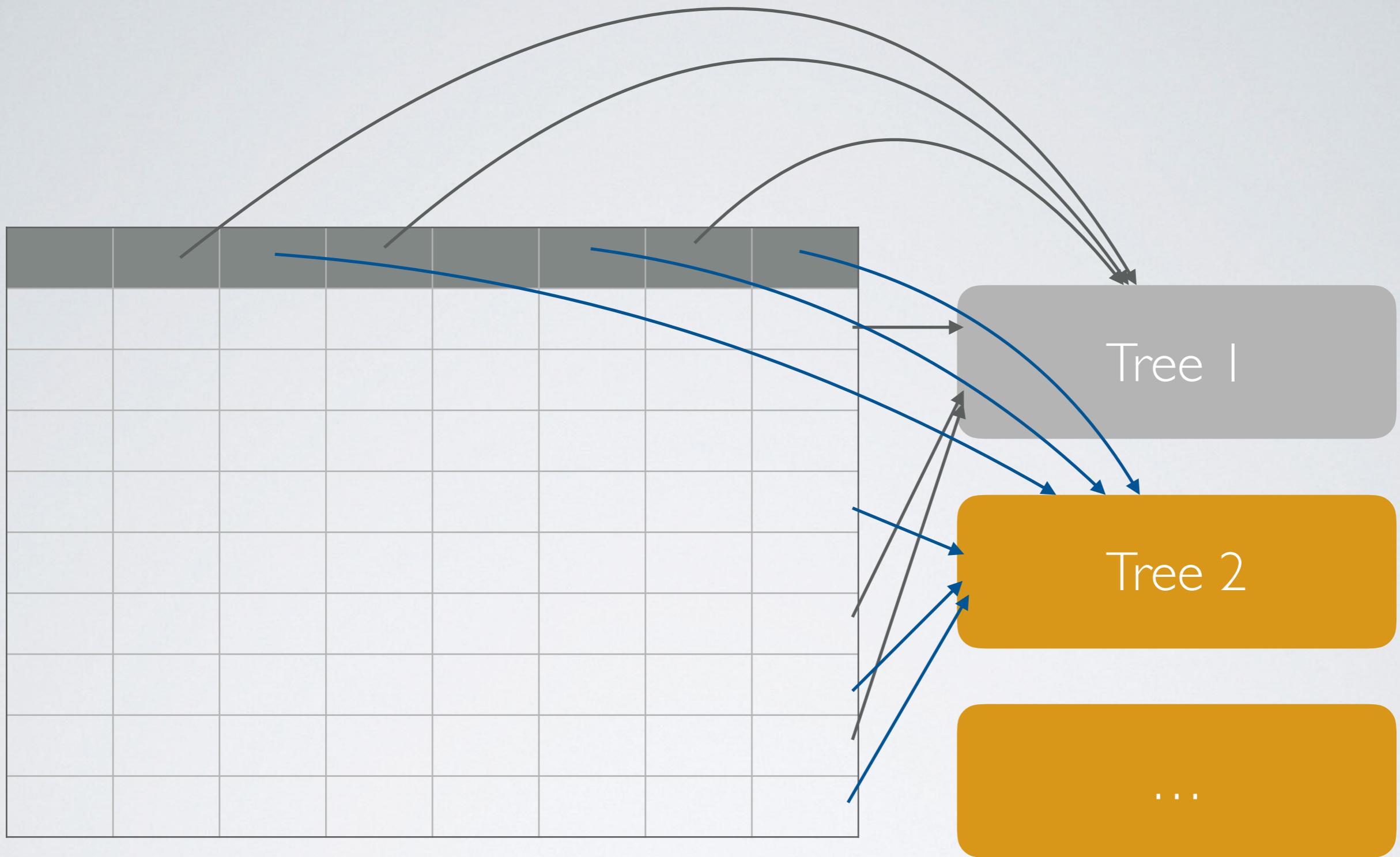
CAN WE DO BETTER?

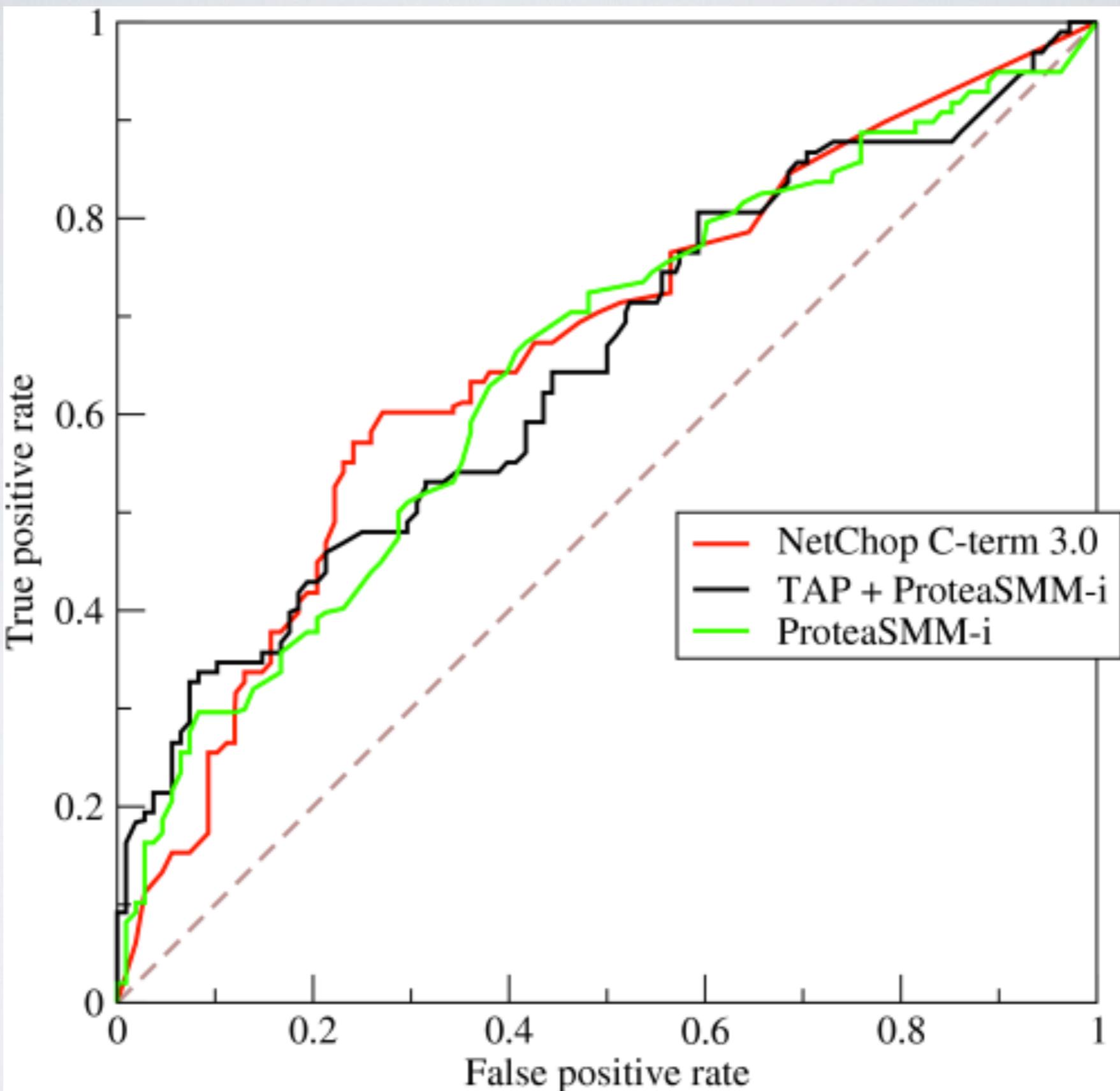
- The model attempts to classify training data, but starts to fail harder with real or test data.
- What about pruning?



RANDOM FORESTS

- Double the random, double the fun!
- We create random decision trees with randomly selected features for branching within the trees.





Example of ROC

BECOMING AN ARBORIST

- R
 - randomForest
 - caret
- Python
 - scikit-learn
- Azure ML

THE CHALLENGE THAT GOT ME STARTED

- Interact with our API
- GET
 - JSON {"blob":<base64>, "targets":<6 ISAs>}
- POST
 - JSON {"target":<your guess>}
 - {"correct":<solved>, "target":<actual>, ...}

A photograph of a dense redwood forest. In the foreground, there's a mix of green ferns and moss-covered logs. A dirt path leads into the forest. The background is filled with tall, thin redwood trees reaching towards a bright sky.

DEMOTIME

STATISTICIANS, LIKE ARTISTS, HAVE THE BAD
HABIT OF FALLING IN LOVE WITH THEIR MODELS.

- George E.P. Box

What applications to InfoSec are you thinking of?

GOING FURTHER

- BlackHat USA 2016 - Crash Course in Data Science for Hackers
<https://www.blackhat.com/us-16/training/crash-course-in-data-science-for-hackers.html>
- Praetorian's Machine Learning Challenge
<https://www.praetorian.com/challenges/machine-learning/>
- Kaggle
<https://www.kaggle.com>
- Should I Trust You? Explaining the Predictions of Any Classifier - Ribeiro/Singh/Guestrin
<http://arxiv.org/pdf/1602.04938.pdf>

EXTRAS

Code From Our Session on Markov Chains

<https://github.com/cattleguard/R-Junk/tree/master/paper%20scissors%20rock%20spock%20lizard>

SPECIAL THANKS

- Photo Attribution
 - Title Slide “Long Distance Relationship” by Andy M CC BY NC License
 - Slide 7 “Toringo Crabapple bonsai 20” by Sage Ross CC License
 - Slide 10 “Redwood Forest” by Byron Hetrick CC BY SA License