Decision Trees & Random Forests

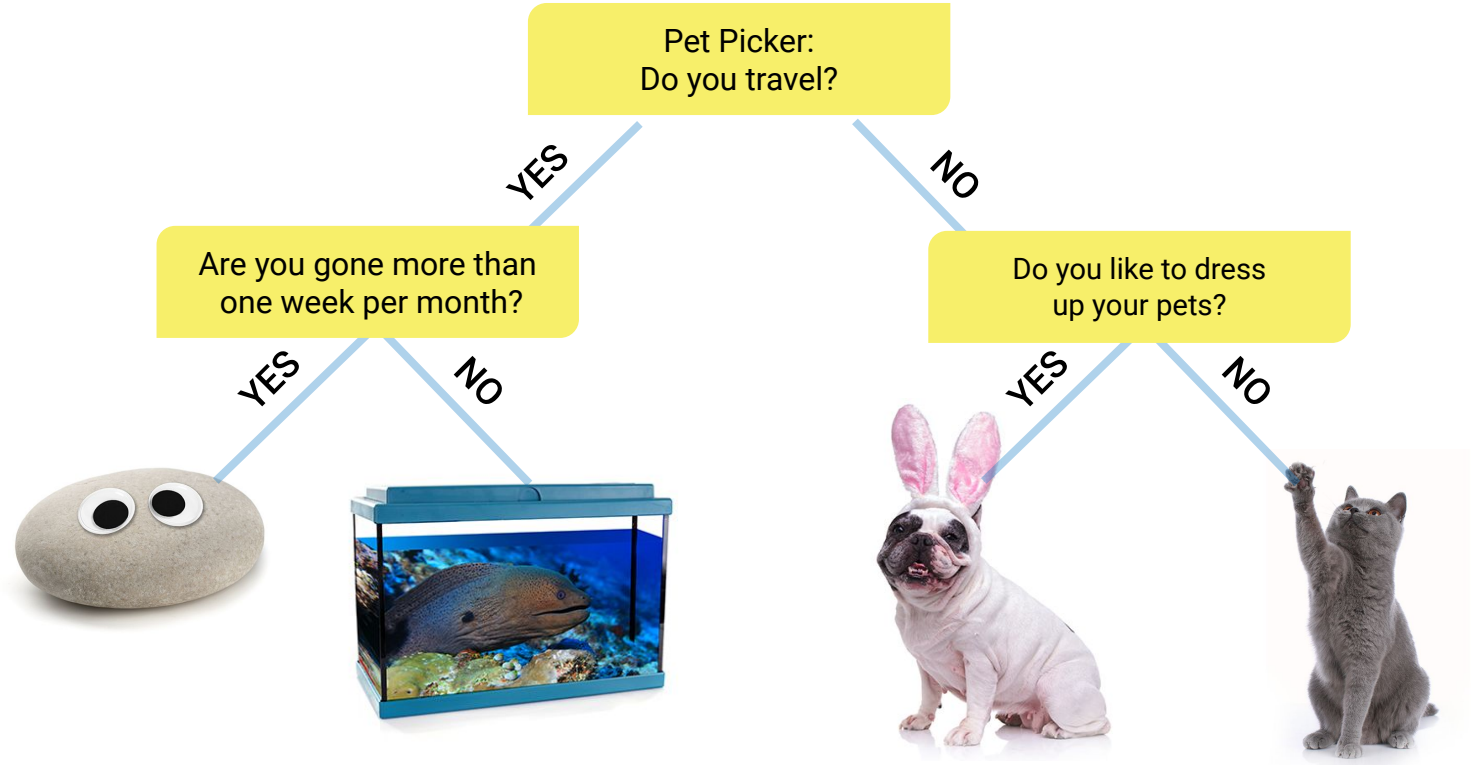
Data Boot Camp

Lesson 21.2

Decision Trees & Random Forests

Decision Trees

Decision trees encode a series of true/false questions.



Decision Trees

These true/false questions can be represented with a series of if/else statements

Do you travel?

Yes Travel:

Are you gone for more than one week per month?

Yes: Pet Rock

No: Pet Fish

No Travel:

Do you like to dress up your pet?

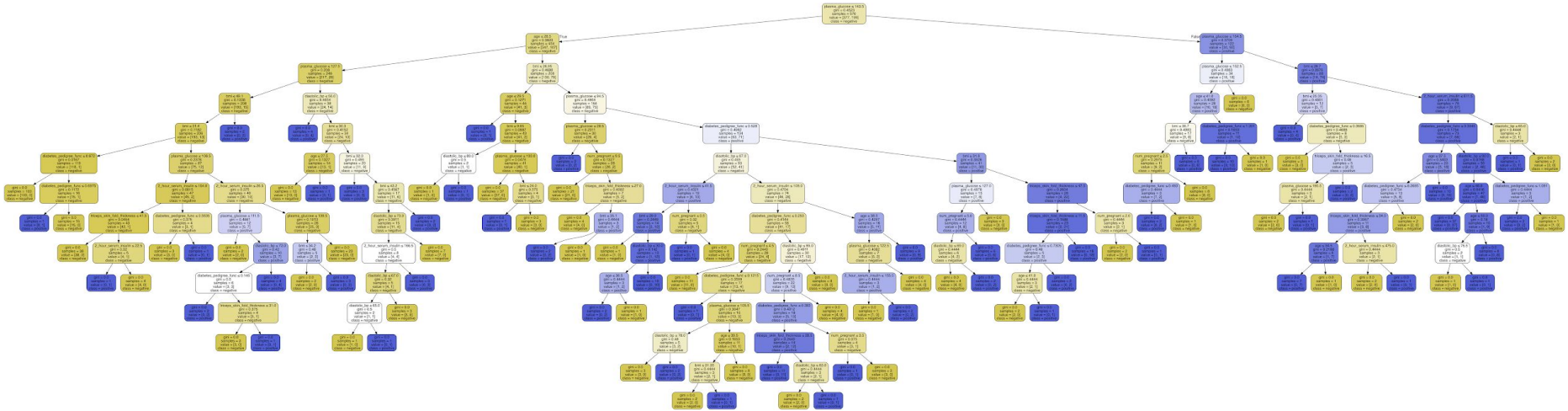
Yes Dress Up: Pet Dog

No Dress Up: Pet Cat

```
if (travel):
  if (time > week):
    print("Rock")
  else:
    print("Fish")
else:
  if (dress_up):
    print("Dog")
  else:
    print("Cat")
```

Decision Tree Complexity

Decision trees can become very complex and may not generalize well.

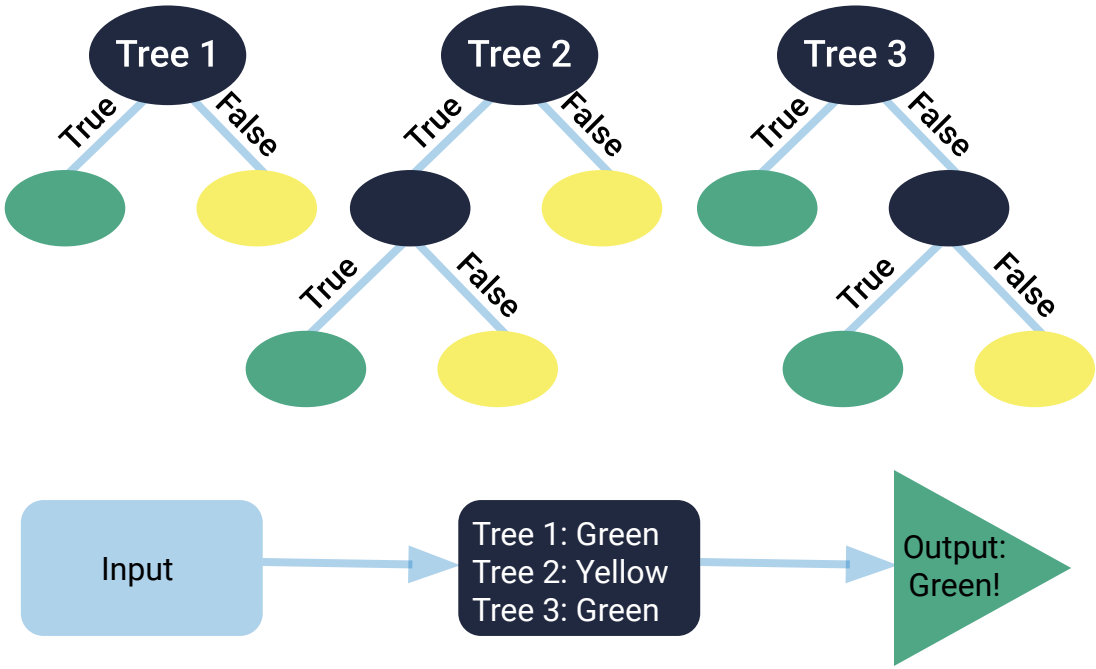


Random Forests

Instead of a single, complex tree like in the previous slide, a random forest algorithm will sample the data and build several smaller, simpler decisions trees (i.e., a forest of trees).

Each tree is much simpler because it is built from a subset of the data.

Each tree is considered a “weak classifier” but when you combine them, they form a “strong classifier.”



Questions?