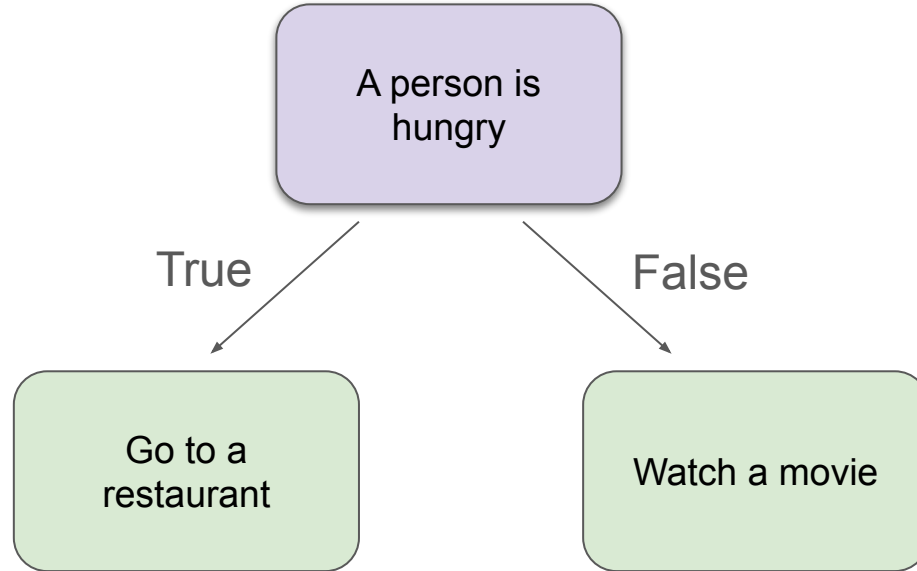


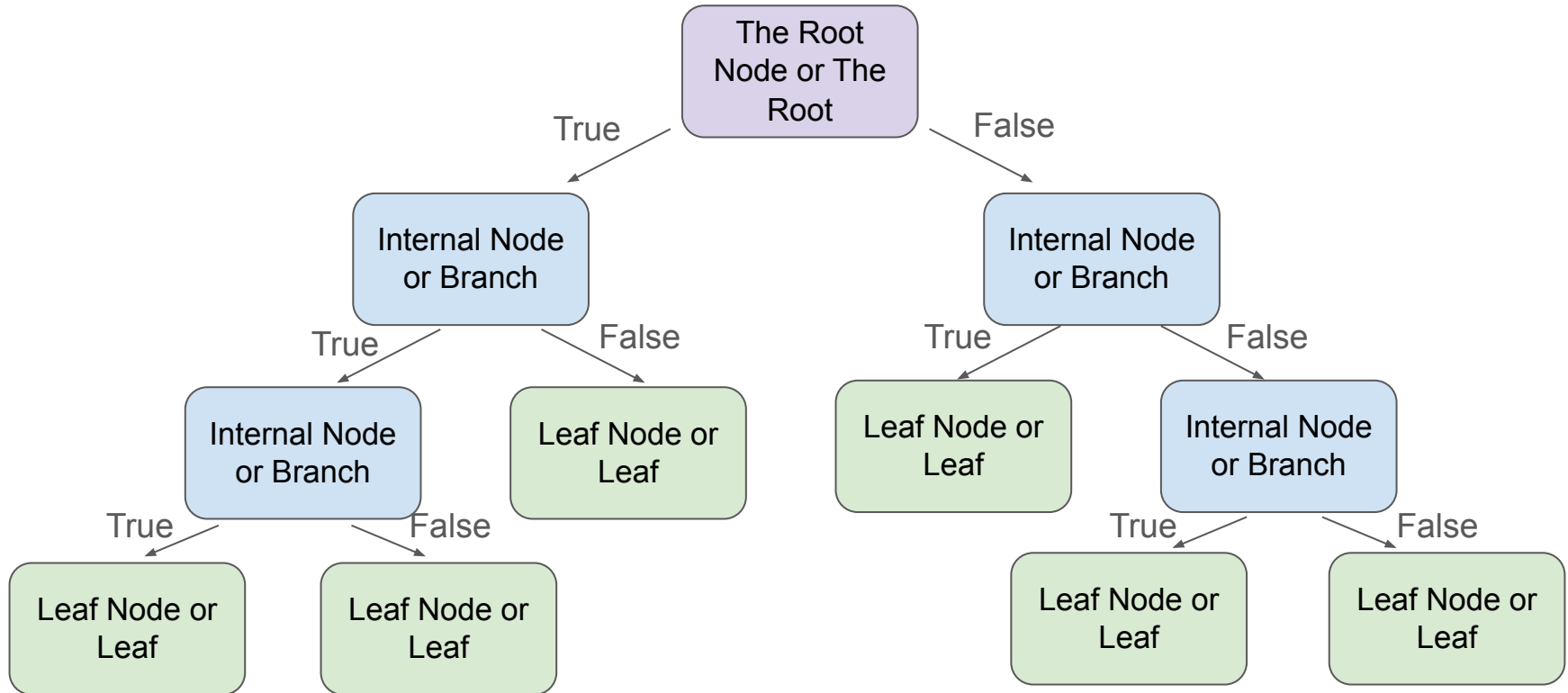
Decision Trees

Presentation by Sarah Vastani

What are decision trees and their use?



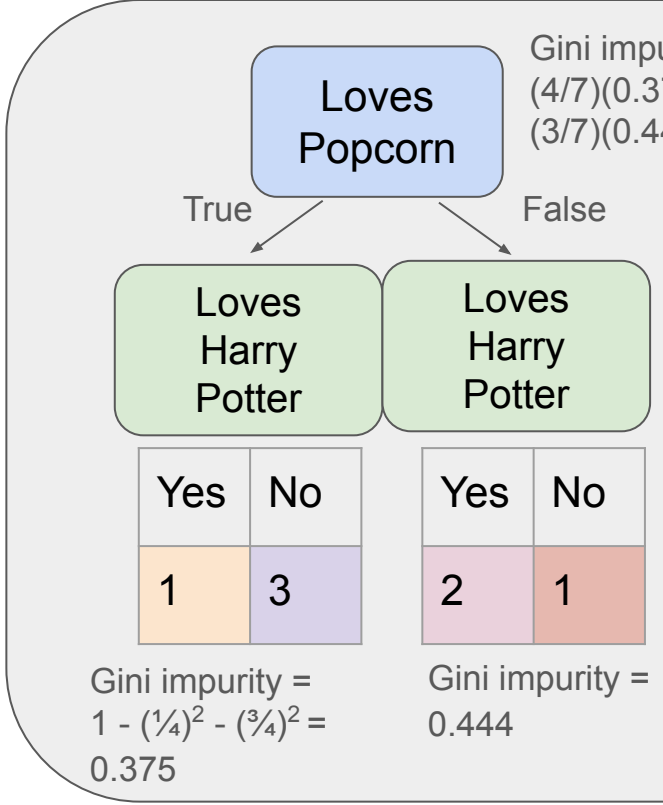
Parts of a decision tree and their interpretation



Example

Loves popcorn	Loves soda	Age	Loves Harry Potter
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No

Yes	Yes	15	???
-----	-----	----	-----



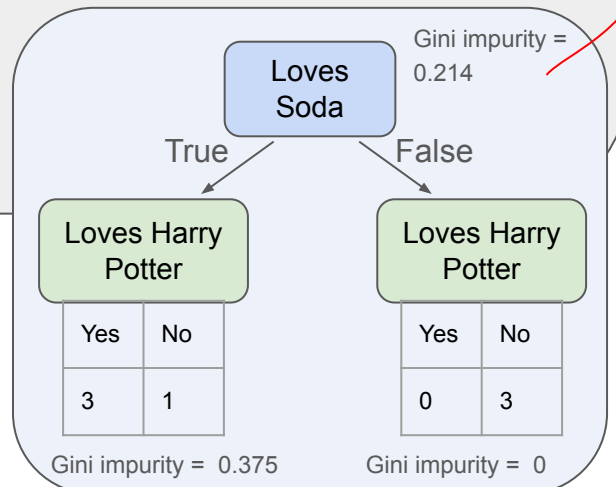
Gini impurity = $(\frac{4}{7})(0.375) + (\frac{3}{7})(0.444) = 0.405$

These are the values that will be compared and used to deduce the root node

Gini impurity for a leaf = $1 - (\text{the probability of "Yes"})^2 - (\text{the probability of "No"})^2$

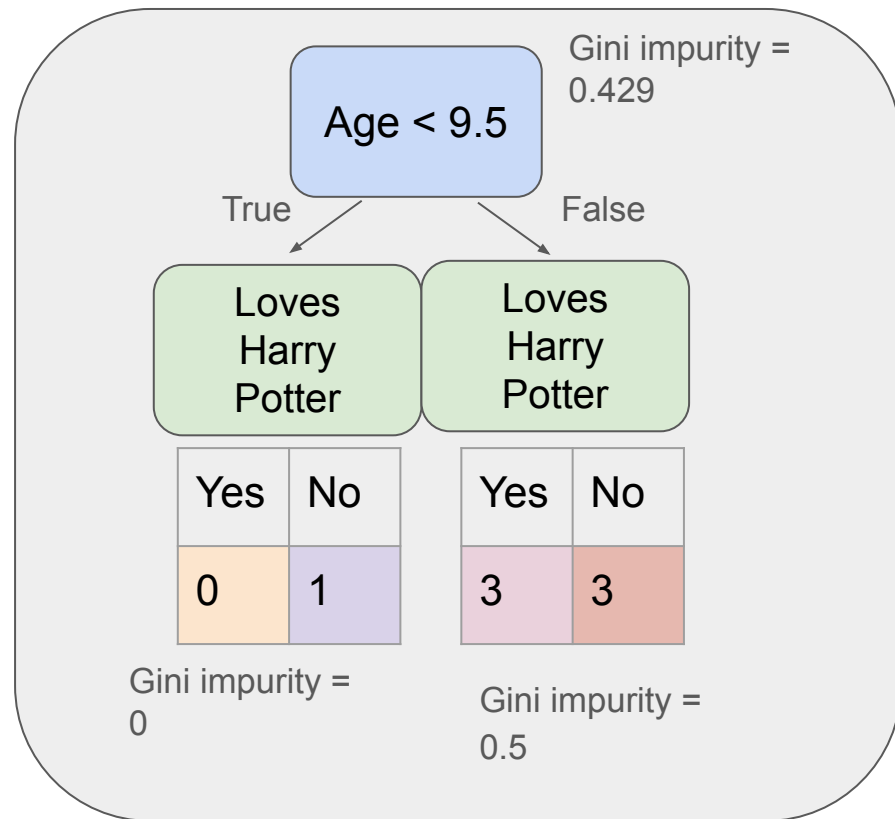
Total Gini impurity = the weighted average of the leaf impurities

Our end goal is to be able to predict this

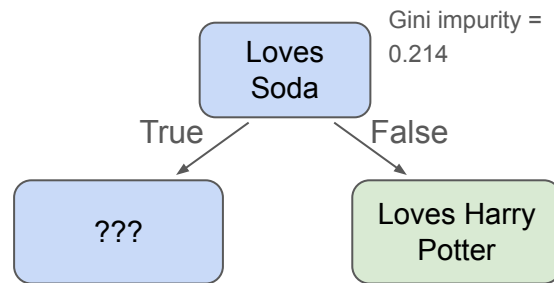
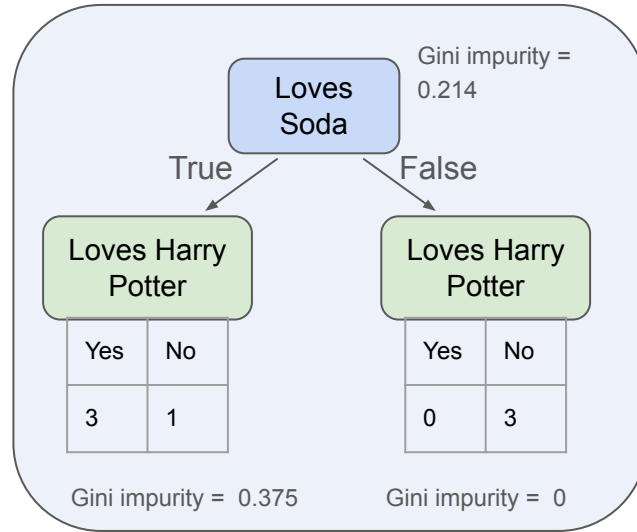


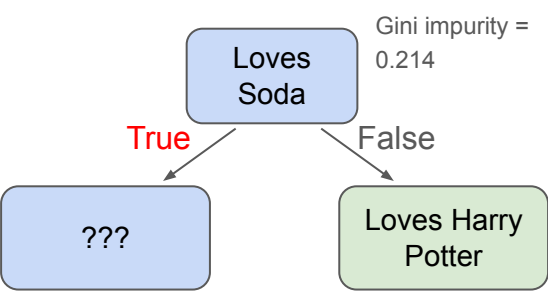
Gini impurity = 0.214

	Age	Loves Harry Potter	
	7	No	
9.5	12	No	Gini impurity = 0.429
15	18	Yes	Gini impurity = 0.343
26.5	35	Yes	Gini impurity = 0.476
36.5	38	Yes	Gini impurity = 0.476
44	50	No	Gini impurity = 0.343
66.5	83	No	Gini impurity = 0.429



Gini impurity for loves popcorn = 0.405
 Gini impurity for loves soda = 0.214 ***
 Gini impurity for age = 0.343

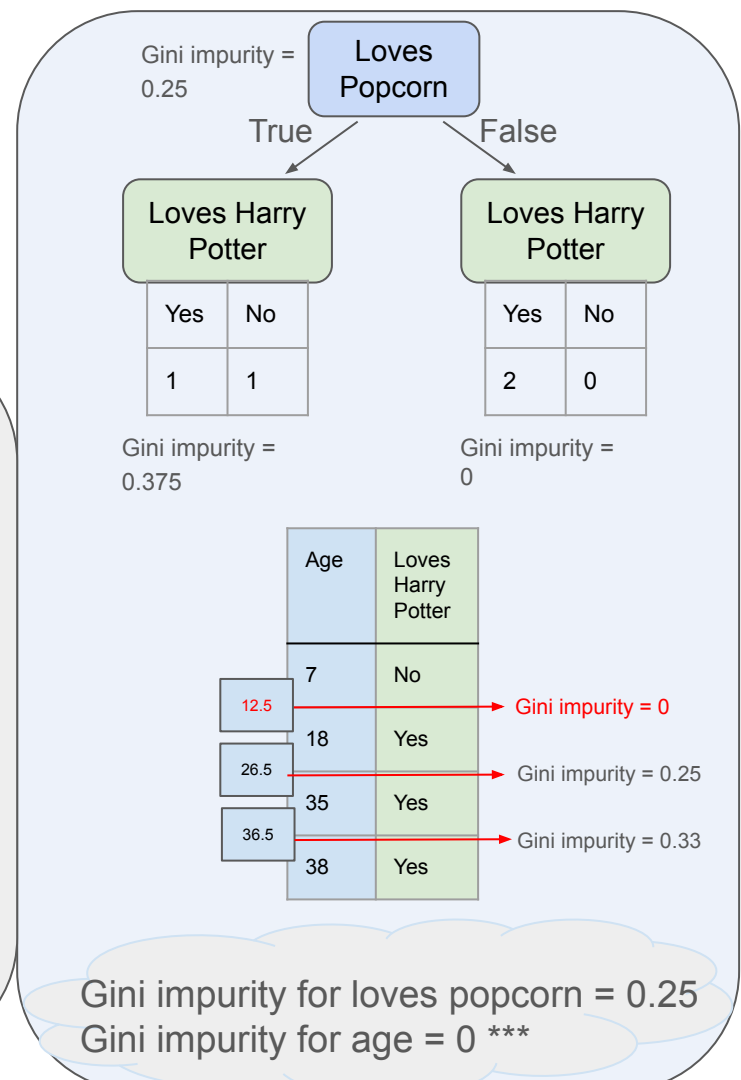




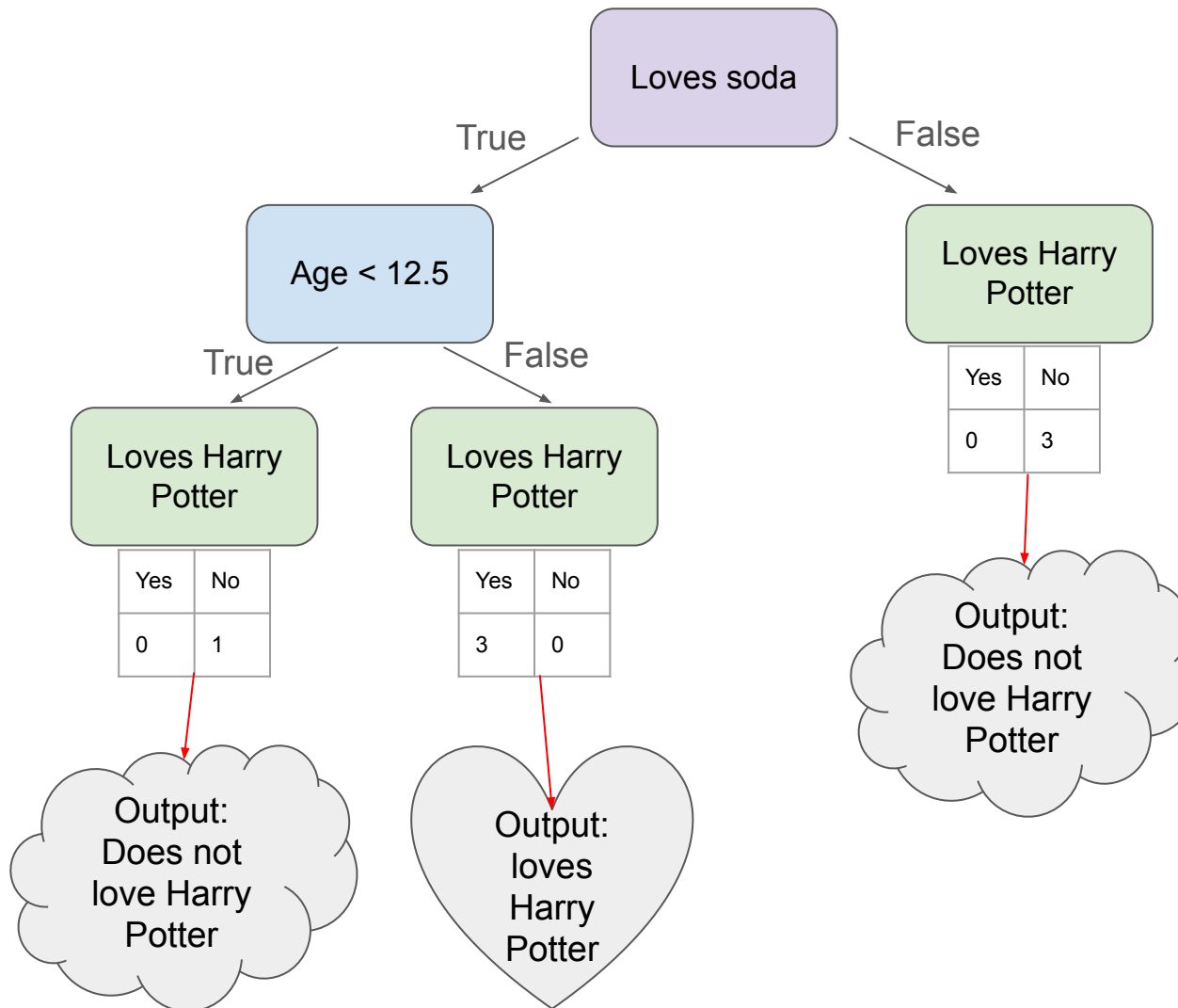
Loves popcorn	Loves soda	Age	Loves Harry Potter
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No

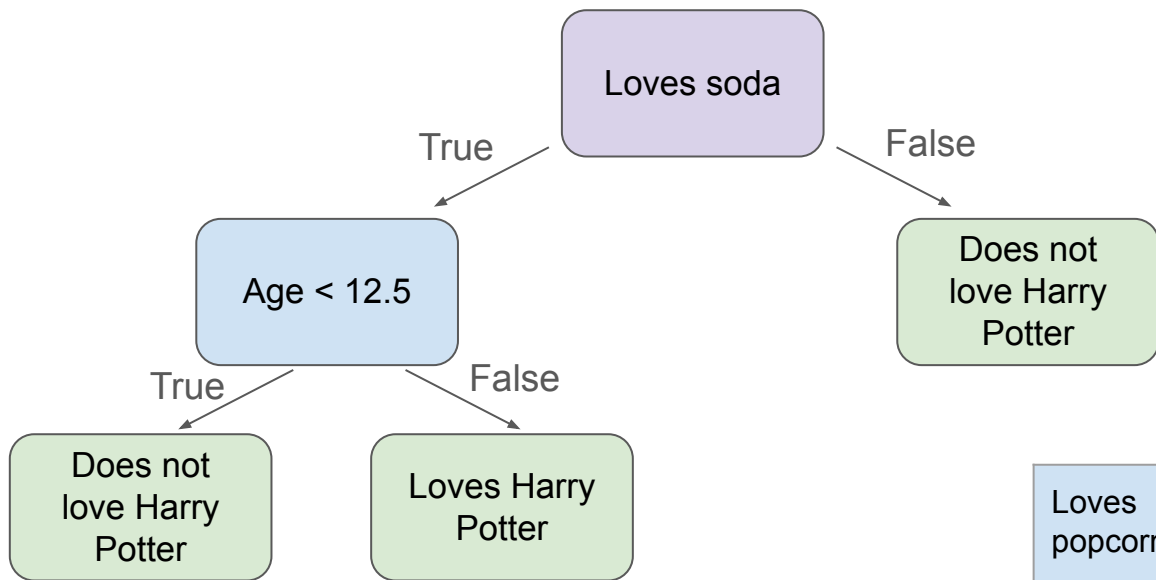


Loves popcorn	Loves soda	Age	Loves Harry Potter
Yes	Yes	7	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes



Loves popcorn	Loves soda	Age	Loves Harry Potter
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No





Yes	No
0	1

Loves popcorn	Loves soda	Age	Loves Harry Potter
Yes	Yes	15	???

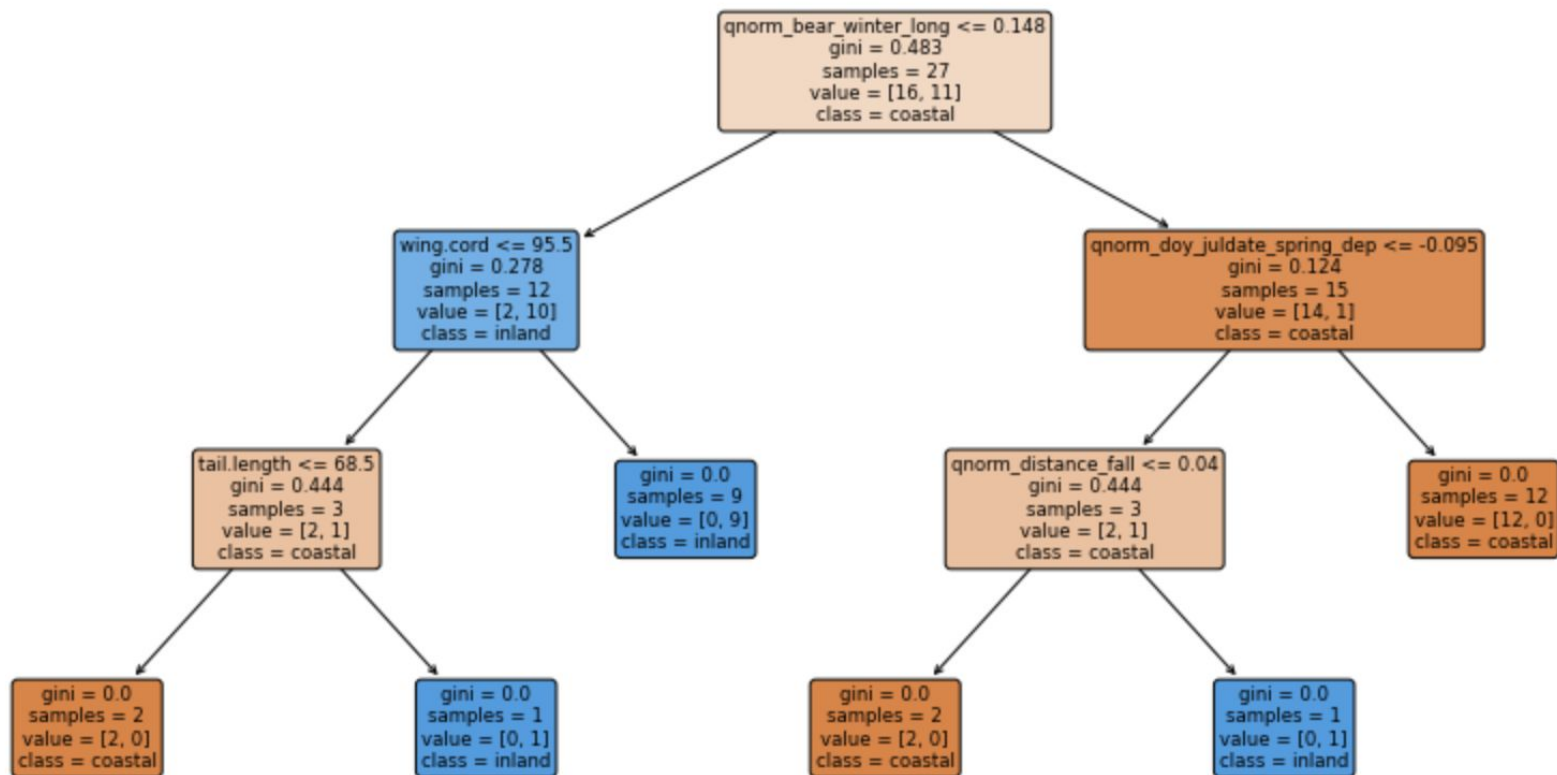


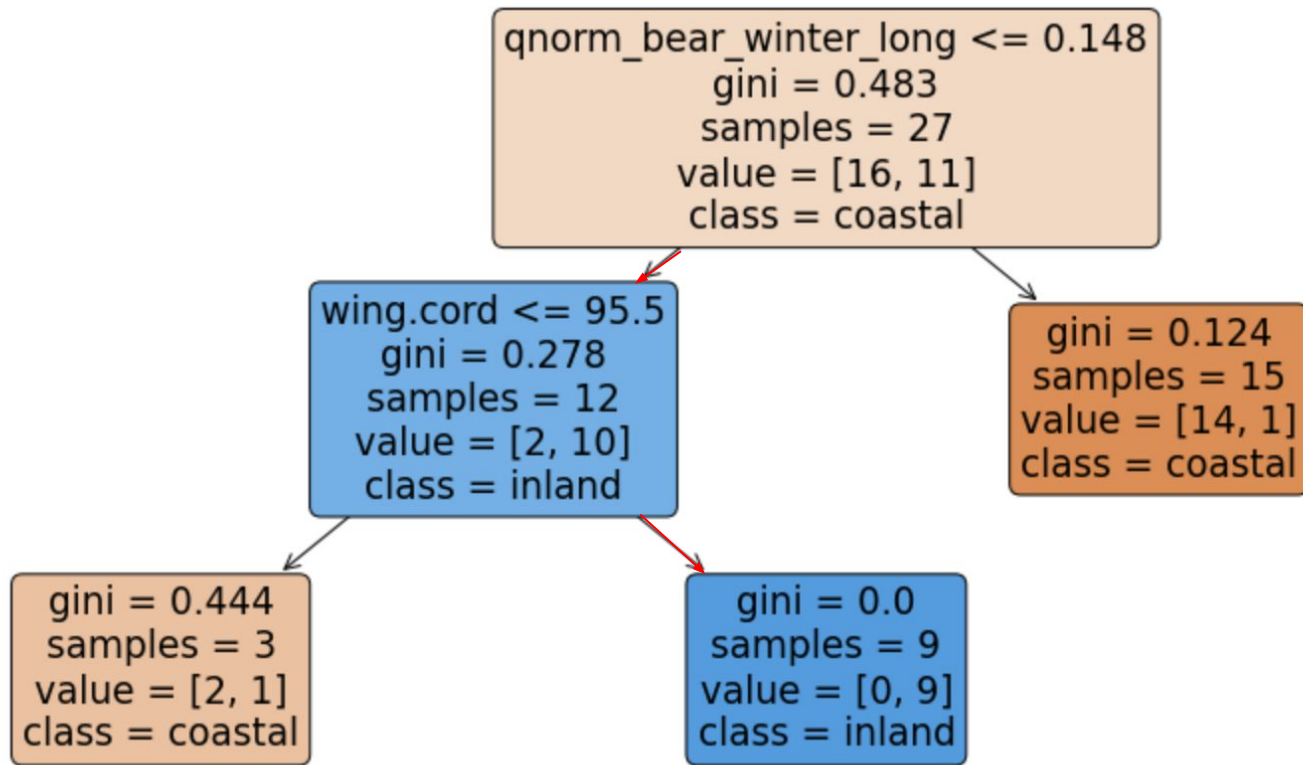
Original Raw Data

speciesID	tarsus. length	fat. score	tail. length	wing. cord	PS_ kipps	p10	winter_ long	fall_ dep	fall_ arr	fall_ duration	spring_ dep	spring_ duration	distance_ fall	distance_ spring
???	29.6	0	78	103.5	32.6	4.9	0.0172	1.034	1.142	0.642	-0.185	0.540	-0.894	-0.765



Data with Imputed Values





speciesID	tarsus. length	fat. score	tail. length	wing. cord	PS_ kipps	p10	winter_ long	fall_ dep	fall_ arr	fall_ duration	spring_ dep	spring_ duration	distance_ fall	distance_ spring
???	29.6	0	78	103.5	32.6	4.9	0.0172	1.034	1.142	0.642	-0.185	0.540	-0.894	-0.765