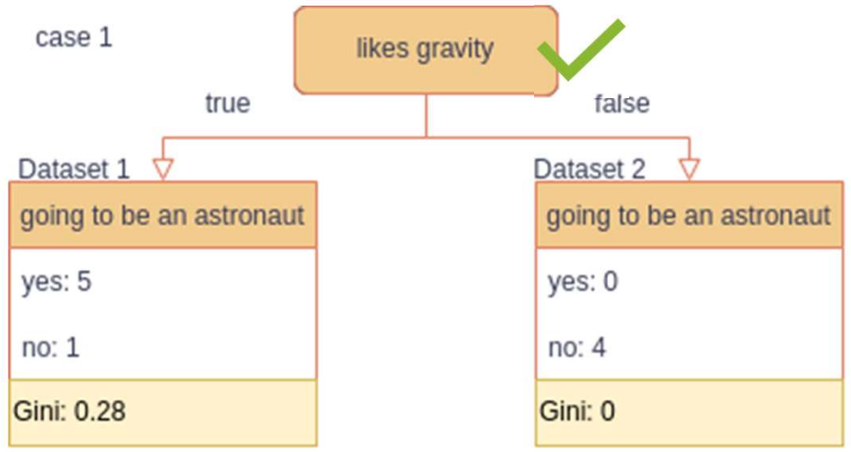
Decision Tree Numeric Examples

Maryam Abdolali

KNTU, Fall 2024

Example 1- Gini

age	likes dogs	likes gravity	going to be an astronaut
24	0	0	0
30	1	1	1
36	0	1	1
36	0	0	0
42	0	0	0
44	1	1	1
46	1	0	0
47	1	1	1
47	0	1	0
51	1	1	1



$$\text{Gini}(D_1) = 1 - \left(\frac{5}{5+1}\right)^2 - \left(\frac{1}{5+1}\right)^2 = 0.28$$

Gini Impurity: 0.168

$$\text{Gini}(D_1) = 1 - \left(\frac{4}{1+4}\right)^2 - \left(\frac{1}{1+4}\right)^2 = 0.32$$

Gini Impurity: 0.32

$$\text{Gini Impurity} = \frac{6}{10} \cdot 0.28 + \frac{4}{10} \cdot 0 = 0.168$$

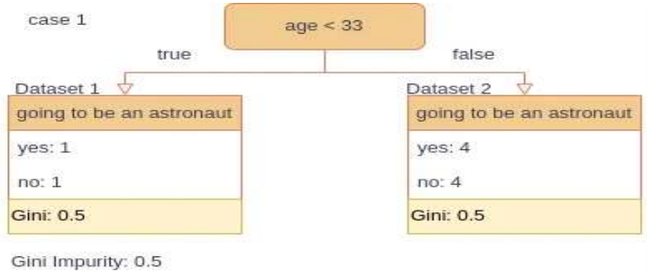
$$\text{Gini Impurity} = \frac{5}{10} \cdot 0.32 + \frac{5}{10} \cdot 0.32 = 0.32$$

https://towardsdatascience.com/decision-trees-for-classification-complete-example-d0bc17fcf1c2

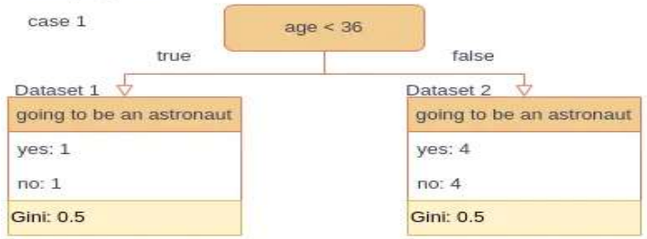
Example 2: Add a numerical Variable

	age	likes dogs	likes gravity	going to be an astronaut
27	24	0	0	0
33	30	1	1	1
36	36	0	1	1
39	36	0	0	0
43	42	0	0	0
45	44	1	1	1
46.5	46	1	0	0
47	47	1	1	1
49	47	0	1	0
	51	1	1	1

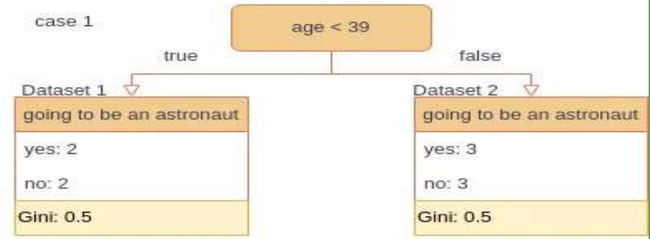
Gini Impurity: 0.441



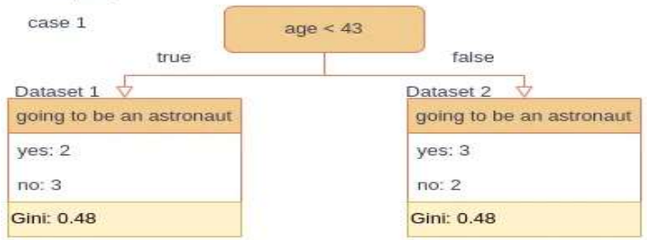
Gini Impurity: 0.5



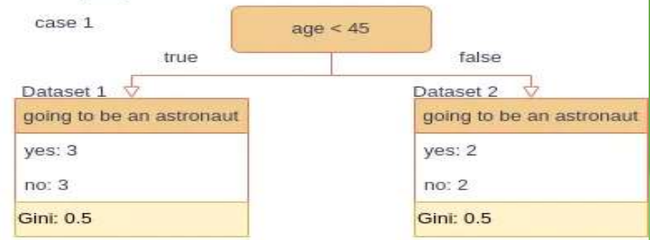
Gini Impurity: 0.5



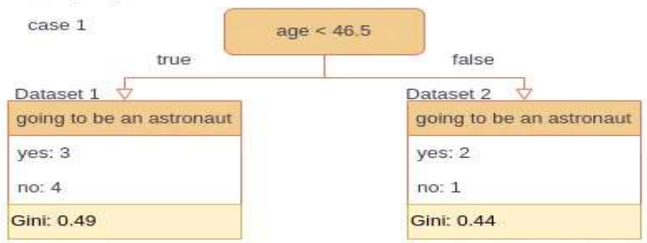
Gini Impurity: 0.5



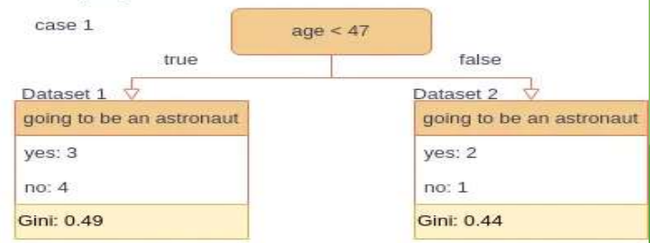
Gini Impurity: 0.48



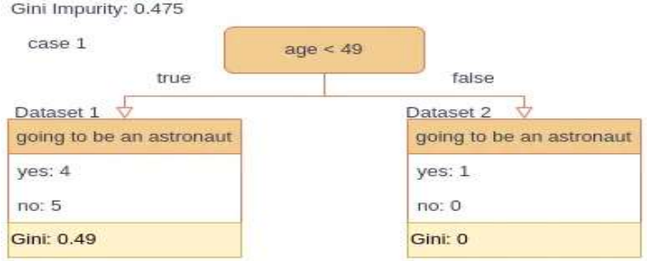
Gini Impurity: 0.5



Gini Impurity: 0.475



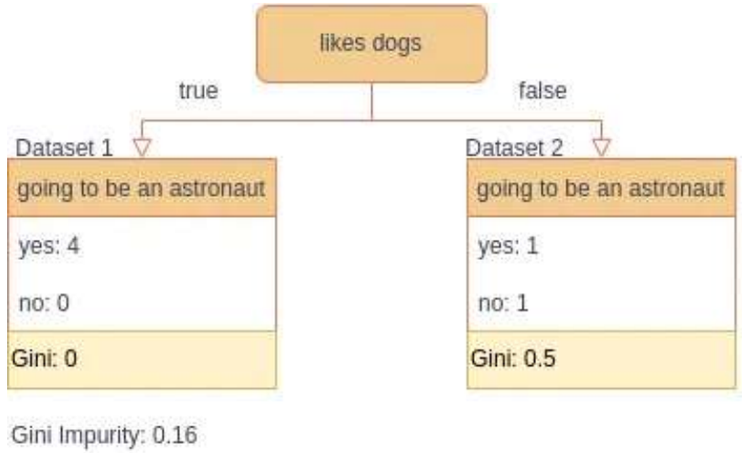
Gini Impurity: 0.475



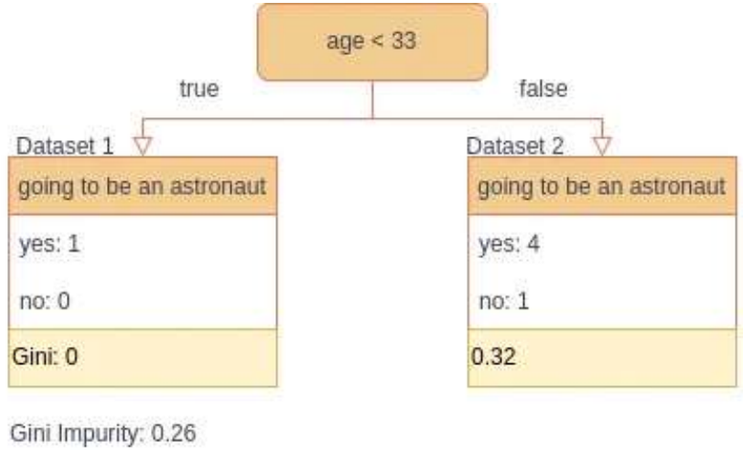
https://towardsdatascience.com/decision-trees-for-classification-complete-example-d0bc17fcf1c2

-cont-

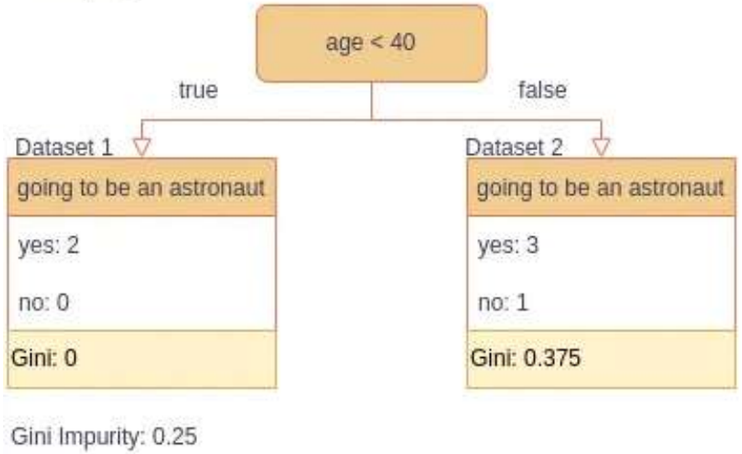
age	likes dogs	going to be an astronaut
30	1	1
36	0	1
44	1	1
47	1	1
47	0	0
51	1	1



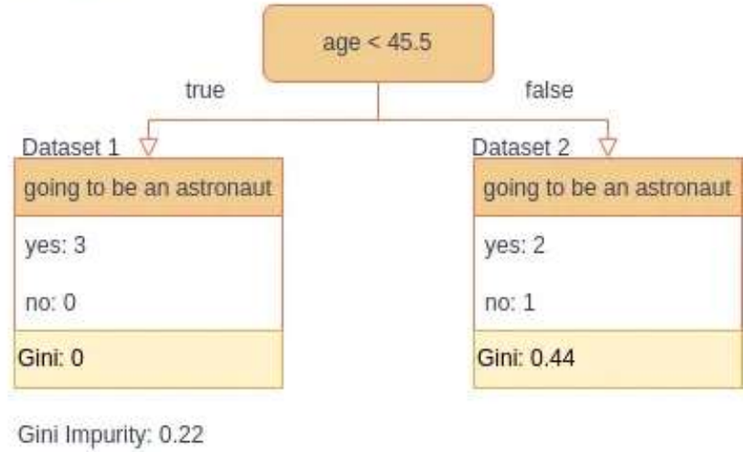
Gini Impurity: 0.16



Gini Impurity: 0.26

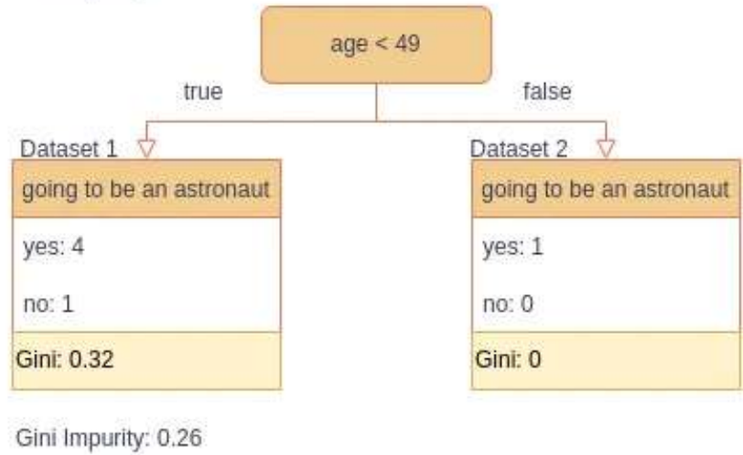


Gini Impurity: 0.25



Gini Impurity: 0.22

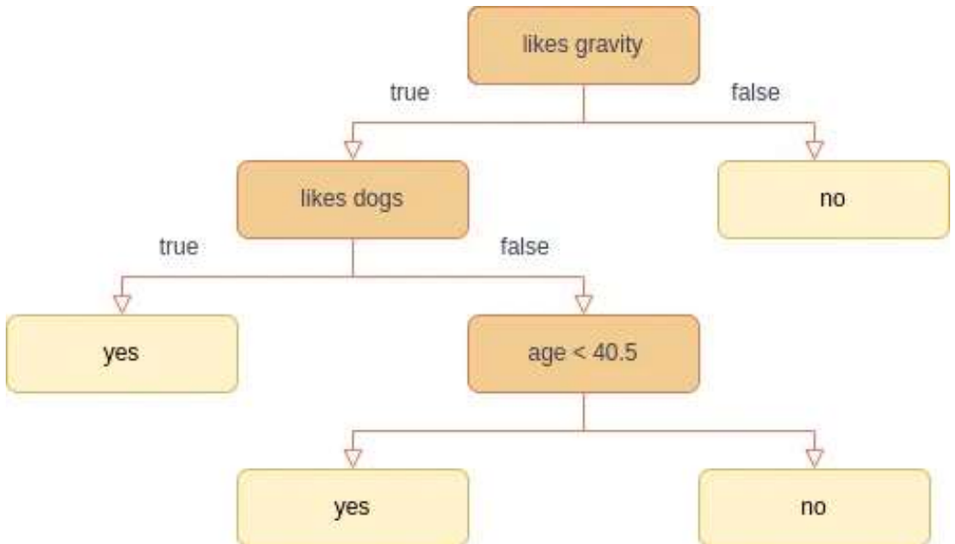
Gini Impurity: 0.22



Gini Impurity: 0.26

https://towardsdatascience.com/decision-trees-for-classification-complete-example-d0bc17fcf1c2

-cont-



https://towardsdatascience.com/decision-trees-for-classification-complete-example-d0bc17fcf1c2

Example 3-

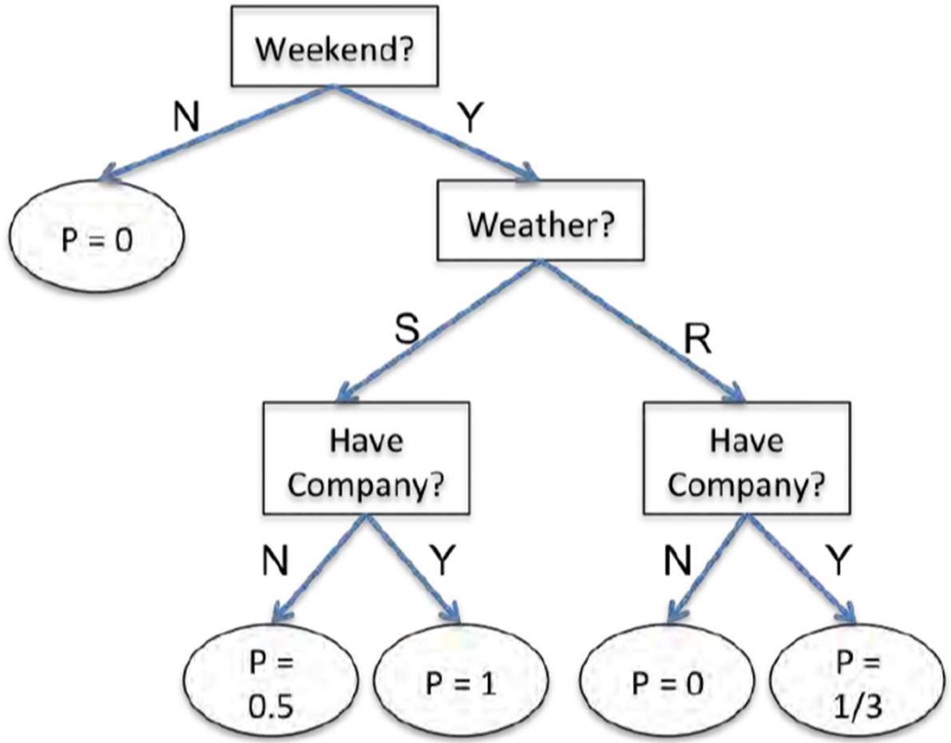
We have some data about when people go hiking. The data take into effect, whether hike is on a weekend or not, if the weather is rainy or sunny, and if the person will have company during the hike.

Weekend?	Company?	Weather	Go Hiking?
Y	N	R	N
Y	Y	R	N
Y	Y	R	Y
Y	Y	S	Y
Y	N	S	Y
Y	N	S	N
Y	Y	R	N
Y	Y	S	Y
N	Y	S	N
N	Y	R	N
N	N	S	N

Solution

- For the first step, the Weekend attribute achieve this:
  - Weekend : $\frac{8}{11}H\left(\frac{1}{2},\frac{1}{2}\right)+\frac{3}{11}H(0,1)=\frac{8}{11}$
  - Weather : $\frac{5}{11}H\left(\frac{1}{5},\frac{4}{5}\right)+\frac{6}{11}H(\frac{1}{2},\frac{1}{2})=\frac{9.6}{11}$
  - Company : $\frac{4}{11}H\left(\frac{1}{4},\frac{3}{4}\right)+\frac{7}{11}H(\frac{3}{7},\frac{4}{7})=\frac{10.1}{11}$
- Therefore we first split on weekend attribute.
  - If weekend = NO: then Go Hiking = NO.
  - If weekend = YES, we need to choose second attribute to split on:
    - Weather : $\frac{4}{8}H\left(\frac{1}{4},\frac{3}{4}\right)+\frac{4}{8}H(\frac{1}{4},\frac{3}{4})=\frac{6.4}{8}$
    - Company : $\frac{5}{8}H\left(\frac{2}{5},\frac{3}{5}\right)+\frac{3}{8}H(\frac{1}{3},\frac{2}{3})=\frac{7.6}{8}$

solution



Q. How about the probability of going to hike on a rainy weekend when having some company?
A. 1/3