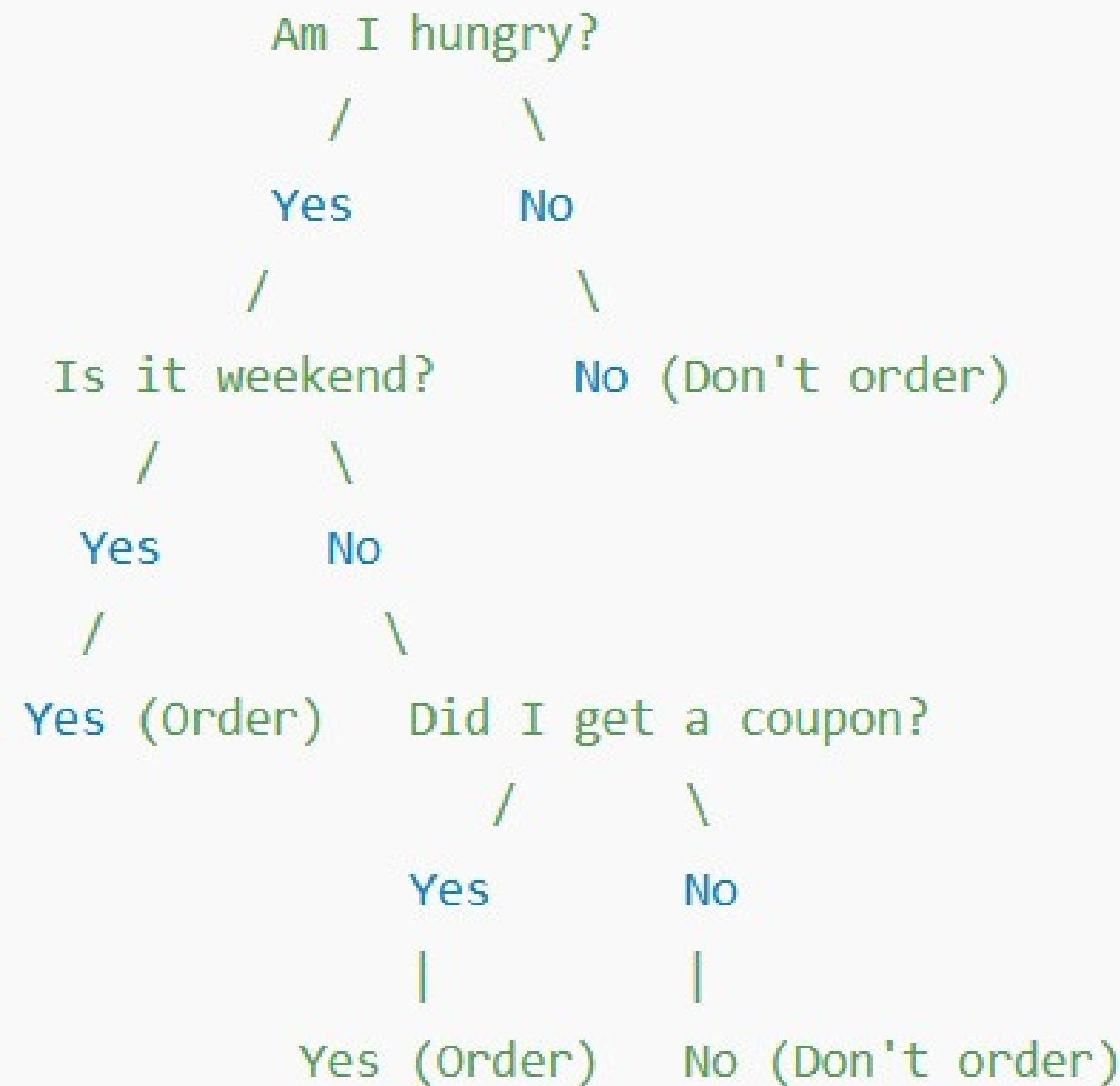


I need to take a decision: Should I order pizza?

weekend	hungry	coupon	Pizza
Yes	Yes	No	Yes
No	Yes	Yes	Yes
No	No	Yes	No
Yes	No	No	No
Yes	Yes	Yes	Yes



GINI AND ENTROPY

Gini Index:

$$\text{Gini} = 1 - \sum(p_i)^2$$

A Before any split — Full Data:

- Total samples = 5
- Yes = 3
- No = 2

$$\text{Gini} = 1 - (3/5)^2 - (2/5)^2 = 1 - 0.36 - 0.16 = 0.48$$

GINI AND ENTROPY

B Split on Am I hungry?

- Yes branch (Hungry = Yes):

- Samples = 3
- Order = Yes (3)
- Gini = $1 - (3/3)^2 = 0$ (pure)

- No branch (Hungry = No):

- Samples = 2
- Order = No (2)
- Gini = $1 - (2/2)^2 = 0$ (pure)

hungry	Pizza
Yes	Yes
Yes	Yes
No	No
No	No
Yes	Yes

🎯 Weighted Gini after split:

$$\text{Gini} = (3/5) \times 0 + (2/5) \times 0 = 0$$

GINI AND ENTROPY

C Split on Is it weekend?

- Yes branch:
 - Samples = 3
 - Yes = 2, No = 1
 - Gini = $1 - (2/3)^2 - (1/3)^2 = 0.444$
- No branch:
 - Samples = 2
 - Yes = 1, No = 1
 - Gini = $1 - (1/2)^2 - (1/2)^2 = 0.5$

weekend	Pizza
Yes	Yes
No	Yes
No	No
Yes	No
Yes	Yes

🎯 Weighted Gini:

$$(3/5) \times 0.444 + (2/5) \times 0.5 = 0.466$$

HOW TO JUDGE GINI OR ENTROPY VALUES?

Purity	Gini	Entropy
Perfectly Pure (All Yes or No)	0	0
Mixed (50-50)	0.5	1
Worse Mix	>0	>0

ENTROPY

$$\text{Entropy} = -p_1 \cdot \log_2(p_1) - p_2 \cdot \log_2(p_2)$$

A Full Dataset Example:

- Total = 5
- Yes = 3 (60%)
- No = 2 (40%)

$$\text{Entropy} = -(0.6) \cdot \log_2(0.6) - (0.4) \cdot \log_2(0.4)$$

Using a calculator:

- $\log_2(0.6) \approx -0.737 \rightarrow 0.6 \times 0.737 = 0.442$
- $\log_2(0.4) \approx -1.322 \rightarrow 0.4 \times 1.322 = 0.529$

$$\text{Entropy} = 0.442 + 0.529 = 0.971$$

ENTROPY

✓ Now we split based on "Am I Hungry?"

1 Hungry = Yes branch:

- Samples: 3
- Labels: Yes, Yes, Yes
- All are Yes → Pure group

Entropy = 0



Weighted Entropy Calculation:

$$\text{Weighted Entropy} = \frac{3}{5} \times 0 + \frac{2}{5} \times 0 = 0$$