Decision Tree

Prof. Surya Prakash
Department of CSE, IIT Indore

Simple Health Dataset: Exercise & Diet vs. Heart Disease

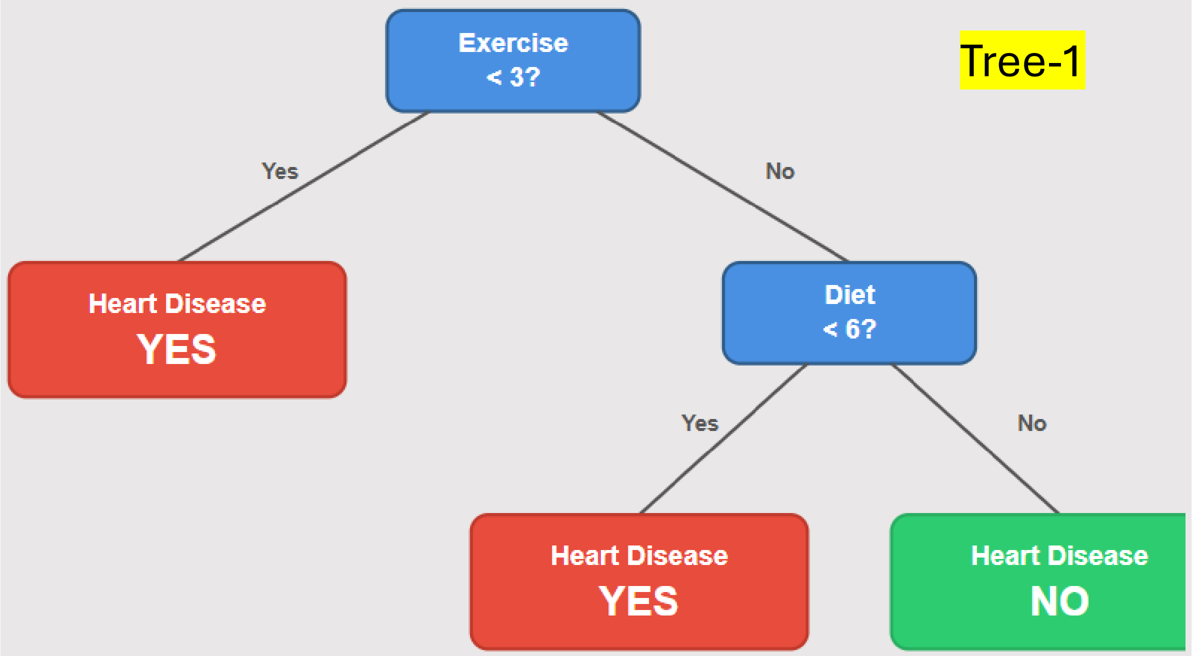
Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
1	0	3	Yes
2	1	4	Yes
3	2	5	Yes
4	3	6	No
5	4	7	No
6	5	8	No
7	3	4	Yes
8	4	6	No

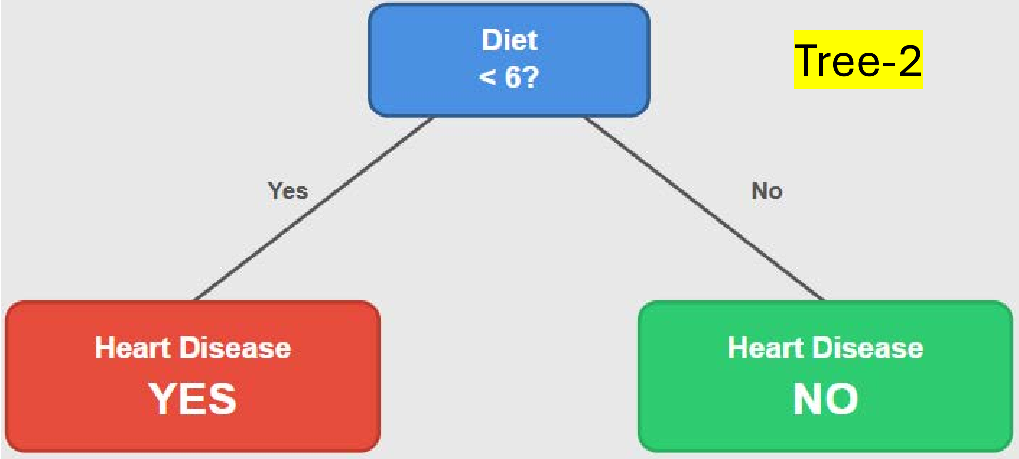
Training Data

Test Person	Exercise (hrs/week)	Diet Quality (1–10)	Predicted Heart Disease
T1	2	5	
T2	4	7	

Test Data

Two Decision Trees – which one is better?





Heart Disease Decision Trees

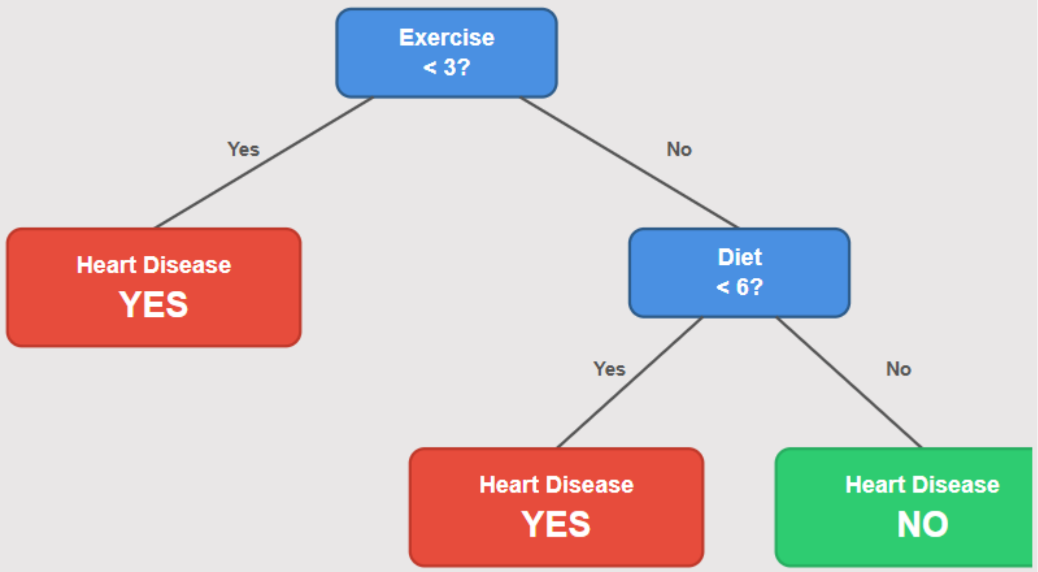
- Both can perform classification, but which one is better?
- Tree-2 is better as it requires less inference time
- Decision tree building algorithm helps to get the **best tree**

Introduction to Decision Trees

- **What is a Decision Tree?**
  - A decision tree is a flowchart-like structure used for decision making or classification.
- **Uses:**
  - Classification
  - Regression
  - Feature selection

Structure of a Decision Tree

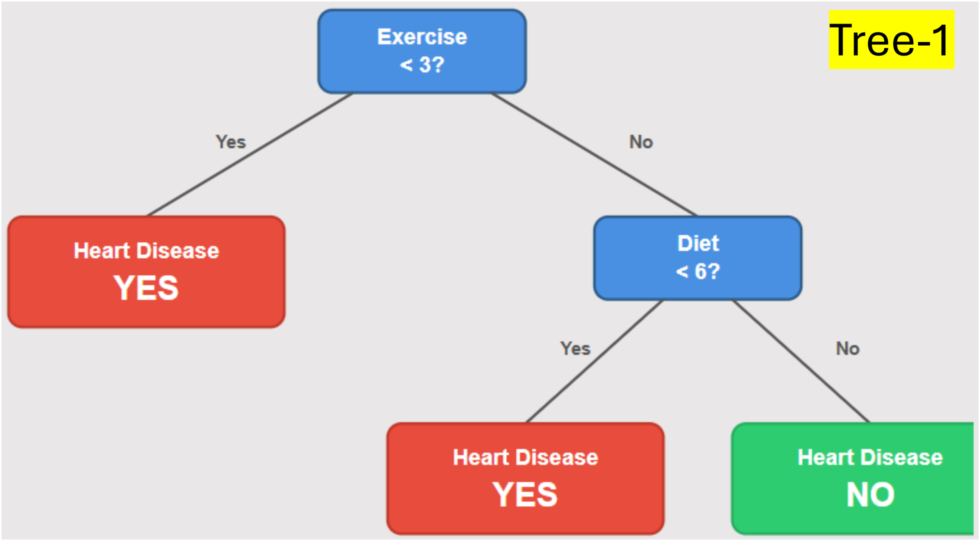
- **Root Node:**
  - Represents the entire dataset and the first decision point
- **Internal Nodes:**
  - Represent decisions based on attributes



- **Leaf Nodes:**
  - Represent the output class/label
- **Edges:**
  - Represent outcomes of the decision at a node

Building a Decision Tree

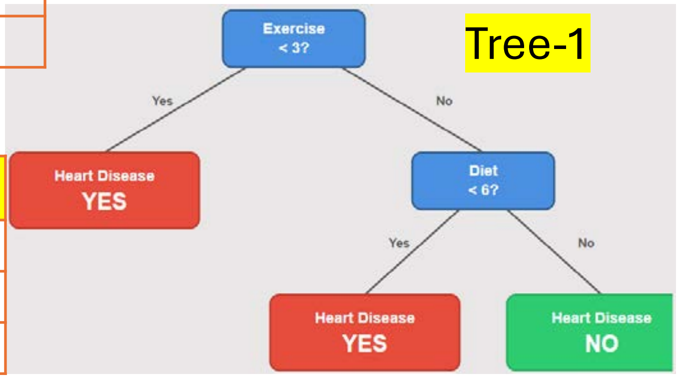
- The **main objective** in building a **Decision Tree** is to **select the best feature (attribute)** at each level (or node) to **split the data** so that the resulting subsets are as **pure** as possible.
- Purity of subsets?
  - This meaning the data points within each subset mostly belong to **one class**.



Building a Decision Tree

Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
1	0	3	Yes
2	1	4	Yes
3	2	5	Yes
4	3	6	No
5	4	7	No
6	5	8	No
7	3	4	Yes
8	4	6	No

Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
4	3	6	No
5	4	7	No
6	5	8	No
7	3	4	Yes
8	4	6	No



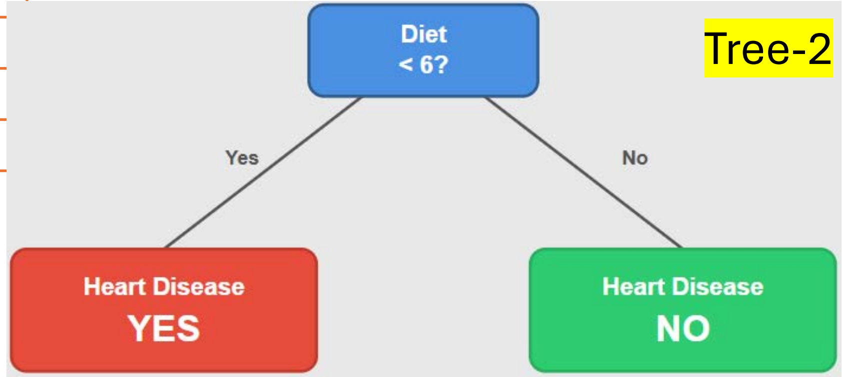
Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
1	0	3	Yes
2	1	4	Yes
3	2	5	Yes

Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
7	3	4	Yes

Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
4	3	6	No
5	4	7	No
6	5	8	No
8	4	6	No

Building a Decision Tree

Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
1	0	3	Yes
2	1	4	Yes
3	2	5	Yes
4	3	6	No
5	4	7	No
6	5	8	No
7	3	4	Yes
8	4	6	No



Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
1	0	3	Yes
2	1	4	Yes
3	2	5	Yes
7	3	4	Yes

Person	Exercise (hrs/week)	Diet Quality (1–10)	Heart Disease
4	3	6	No
5	4	7	No
6	5	8	No
8	4	6	No

Building a Decision Tree

- **ID3: Iterative Dichotomiser 3:**
  - ID3 algorithm is specifically designed for **classification**, not regression.
  - The earlier versions (ID1 and ID2) were unpublished or internal prototypes.
  - ID3 was the first version to be formally published and widely adopted.

ID3 Algorithm

- Uses the concept of
  - Entropy
  - Information gain

ID3 Algorithm

- **Entropy:**
  - In the context of **information theory and decision trees, entropy** is a measure of **impurity or uncertainty** in a dataset.

If a dataset S has c classes, and the probability of class i is p_i, then

the **entropy** of S is:

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i)$$

- **Intuition**
  - If all examples belong to a **single class** → entropy = **0** (pure, no uncertainty).
  - If examples are **evenly split** among classes → entropy is **maximum** (maximum uncertainty).

Consider Dataset - 14 samples

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Example

Suppose we have a dataset of 14 instances of "Play Tennis":

- 9 "Yes"
- 5 "No"

So:

$$p(Yes) = \frac{9}{14}, \quad p(No) = \frac{5}{14}$$

$$Entropy(S) = -\left(\tfrac{9}{14}\log_2 \tfrac{9}{14} + \tfrac{5}{14}\log_2 \tfrac{5}{14}\right) \approx 0.94$$

Information Gain

- **Information Gain** measures how much **uncertainty (entropy) is reduced** when we split a dataset on an attribute.
- Mathematically:

$$IG(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Where:

- S = original dataset
- A = attribute we split on
- $Values(A)$ = all possible values of attribute A
- S_v = subset of S where $A = v$
- $\frac{|S_v|}{|S|}$ = weight (proportion of samples)

Information Gain

- **Intuition**
  - **High IG** → attribute gives a “clearer split” (reduces randomness a lot).
  - **Low IG** → attribute doesn’t help much in classification.

- Decision Trees (like ID3) choose the **attribute with maximum IG** at each node.

ID3 Algorithm

- **Steps:**
  - Calculate Entropy of the dataset
  - Compute Information Gain for each attribute
  - Select attribute with highest Information Gain as the decision node
  - Repeat recursively for each child node until stopping criteria is met

When to Stop Splitting

- All samples belong to one class
- No remaining attributes
- Maximum depth reached

Entropy of the full dataset S

Total instances = 14

`Play = yes` : $9 \rightarrow p_{yes} = \frac{9}{14}$

`Play = no` : $5 \rightarrow p_{no} = \frac{5}{14}$

$$Entropy(S) = -p_{yes}\log_2(p_{yes}) - p_{no}\log_2(p_{no})$$

$$= -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$$

$$= -0.643 \cdot \log_2(0.643) - 0.357 \cdot \log_2(0.357) \approx 0.940$$

Step 2: Information Gain for each attribute

- How many attributes ? Four

- **Four attributes:**
  - Outlook
  - Temperature
  - Humidity
  - Windy

- Compute information Gain *w.r.t.* each attribute, and choose the attribute for splitting with maximum information gain

Step 2: Information Gain for each attribute

- **Attribute: Outlook**
  - Values: **sunny**, **overcast, rainy**

- **sunny**: 5 samples (Play = 2 yes, 3 no)

$$p_{yes} = \frac{2}{5}, \quad p_{no} = \frac{3}{5}$$

$$Entropy = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) \approx 0.971$$

Step 2: Information Gain for each attribute

- **Attribute : Outlook**
  - Values: **sunny**, **overcast**, **rainy**

- **overcast**: 4 samples (All yes)

$$
\begin{aligned}
p_{yes} &= 1, \quad p_{no} = 0 \\
Entropy &= -p_{yes} \cdot \log_2(p_{yes}) - p_{no} \cdot \log_2(p_{no}) \\
&= -1 \cdot \log_2(1) - 0 \cdot \log_2(0) \\
&= -1 \cdot 0 - 0 = \boxed{0}
\end{aligned}
$$

Step 2: Information Gain for each attribute

- **Attribute : Outlook**
  - Values: **sunny**, **overcast**, **rainy**

- **rainy**: 5 samples (Play = 3 yes, 2 no)

$$p_{yes} = \frac{3}{5}, \quad p_{no} = \frac{2}{5}$$

$$\begin{aligned} Entropy &= -\frac{3}{5} \cdot \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \cdot \log_2\left(\frac{2}{5}\right) \\ &= -0.6 \cdot (-0.737) - 0.4 \cdot (-1.322) \\ &= 0.442 + 0.529 \\ &= \boxed{0.971} \end{aligned}$$

Step 2: Information Gain for each attribute

- **Information gain w.r.t Outlook**

$$Information\ Gain(S, Attribute) = Entropy(\text{Parent}) - Weighted\ Average\ Entropy\ (Children)$$

- **For Attribute = Outlook**

$$Information\ Gain(S, Outlook) = Entropy(S) - \left(\frac{5}{14} \times Entropy(\text{Sunny}) + \frac{4}{14} \times Entropy(\text{Overcast}) + \frac{5}{14} \times Entropy(\text{Rainy})\right)$$

$$Gain(S, Outlook) = 0.940 - \left(\frac{5}{14} \cdot 0.971 + \frac{4}{14} \cdot 0 + \frac{5}{14} \cdot 0.971\right)$$

$$= 0.940 - (0.347 + 0 + 0.347) = 0.940 - 0.694 = \boxed{0.246}$$

Outlook	Count	Entropy	Weight
sunny	**5**	0.971	**5/14**
sunny	**4**	**0**	**4/14**
overcast	**5**	0.971	**5/14**

Step 2: Information Gain for each attribute

- **Attribute: Temperature**
  - **Values: hot, mild, cool**

- **hot**: 4 samples (2 yes, 2 no)

$$p_{yes} = \frac{2}{4}, \quad p_{no} = \frac{2}{4}$$

$$Entropy = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right) = -2 \cdot \left(\frac{1}{2} \cdot \log_2\frac{1}{2}\right) = 1.0$$

Step 2: Information Gain for each attribute

- **Attribute: Temperature**
  - **Values: hot, mild, cool**

- **mild:** 6 samples (4 yes, 2 no)

$$p_{yes} = \frac{4}{6}, \quad p_{no} = \frac{2}{6}$$

$$Entropy = -\left(\frac{4}{6}\log_2\frac{4}{6} + \frac{2}{6}\log_2\frac{2}{6}\right) \approx -(0.667 \cdot (-0.585) + 0.333 \cdot (-1.585)) \approx 0.918$$

Step 2: Information Gain for each attribute

- **Attribute: Temperature**
  - **Values: hot, mild, cool**

- **cool :** 4 samples (3 yes, 1 no)

$$p_{yes} = \frac{3}{4}, \quad p_{no} = \frac{1}{4}$$

$$Entropy = -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) = -(0.75 \cdot (-0.415) + 0.25 \cdot (-2)) \approx 0.811$$

Step 2: Information Gain for each attribute

- **Information gain w.r.t Temperature**

$$Information\ Gain(S, Attribute) = Entropy(\text{Parent}) - Weighted\ Average\ Entropy\ (Children)$$

- **For Attribute = Temperature**

Temperature	Count	Entropy	Weight
hot	**4**	1	**4/14**
mild	**6**	**0.918**	**6/14**
cool	**4**	0.811	**4/14**

$$Information\ Gain(S, temperature) = Entropy(S) - \left(\frac{4}{14} \times Entropy(hot) + \frac{6}{14} \times Entropy(mild) + \frac{4}{14} \times Entropy(cool)\right)$$

$$= 0.940 - \left(\frac{4}{14} \cdot 1.0 + \frac{6}{14} \cdot 0.918 + \frac{4}{14} \cdot 0.811\right)$$

$$= 0.940 - (0.286 + 0.393 + 0.232) = 0.940 - 0.911 = \boxed{0.029}$$

Step 2: Information Gain for each attribute

- **Attribute: Humidity**
  - **Values: high, normal**

- **high**: 7 samples (3 yes, 4 no)

$$p_{yes} = \frac{3}{7}, \quad p_{no} = \frac{4}{7}$$

$$Entropy(high) = -\left(\frac{3}{7}\log_2\frac{3}{7} + \frac{4}{7}\log_2\frac{4}{7}\right)$$

$$= -(0.429 \cdot (-1.222) + 0.571 \cdot (-0.807)) = 0.985$$

- **normal**: 7 samples (6 yes, 1 no)

$$p_{yes} = \frac{6}{7}, \quad p_{no} = \frac{1}{7}$$

$$Entropy(normal) = -\left(\frac{6}{7}\log_2\frac{6}{7} + \frac{1}{7}\log_2\frac{1}{7}\right)$$

$$= -(0.857 \cdot (-0.222) + 0.143 \cdot (-2.807)) = 0.592$$

Step 2: Information Gain for each attribute

- **Information gain w.r.t Humidity**

$$Information\ Gain(S, Attribute) = Entropy(\text{Parent}) - Weighted\ Average\ Entropy\ (Children)$$

- **For Attribute = Humidity**

Humidity	Count	Entropy	Weight
high	7	0.985	7/14
normal	7	0.592	7/14

$$Information\ Gain(S, humidity) = Entropy(S) - \left(\frac{7}{14} \cdot Entropy(high) + \frac{7}{14} \cdot Entropy(normal)\right)$$

$$= 0.940 - \left(\frac{7}{14} \cdot 0.985 + \frac{7}{14} \cdot 0.592\right) = 0.940 - (0.492 + 0.296) = 0.940 - 0.788 = \boxed{0.152}$$

Step 2: Information Gain for each attribute

- **Attribute: Windy**
  - **Values: true, false**

- **true**: 6 samples (3 yes, 3 no)

$$p_{yes} = \frac{3}{6}, \quad p_{no} = \frac{3}{6}$$

$$Entropy(true) = -\left(\frac{3}{6}\log_2\frac{3}{6} + \frac{3}{6}\log_2\frac{3}{6}\right)$$

$$= -\left(0.5 \cdot \log_2 0.5 + 0.5 \cdot \log_2 0.5\right) = -\left(2 \cdot 0.5 \cdot (-1)\right) = 1.0$$

Step 2: Information Gain for each attribute

- **Attribute: Windy**
  - **Values: true, false**

- **false**: 8 samples (6 yes, 2 no)

$$p_{yes} = \frac{6}{8}, \quad p_{no} = \frac{2}{8}$$

$$Entropy(false) = -\left(\frac{6}{8}\log_2\frac{6}{8} + \frac{2}{8}\log_2\frac{2}{8}\right)$$

$$= -(0.75 \cdot \log_2 0.75 + 0.25 \cdot \log_2 0.25)$$

$$= -(0.75 \cdot (-0.415) + 0.25 \cdot (-2)) = 0.811$$

Step 2: Information Gain for each attribute

- **Information gain w.r.t Windy**

$$Information\ Gain(S, Attribute) = Entropy(\text{Parent}) - Weighted\ Average\ Entropy\ (Children)$$

- **For Attribute = Windy**

$$Information\ Gain(S, windy) = Entropy(S) - \left(\frac{8}{14} \cdot Entropy(false) + \frac{6}{14} \cdot Entropy(true)\right)$$

$$= 0.940 - \left(\frac{8}{14} \cdot 0.811 + \frac{6}{14} \cdot 1.0\right)$$

$$= 0.940 - (0.463 + 0.429) = 0.940 - 0.892 = \boxed{0.048}$$

Windy	Count	Entropy	Weight
true	6	1.0	6/14
false	8	0.811	8/14

Information Gain for different attributes – Summary

Feature	Information Gain
Outlook	**0.247**
Humidity	0.152
Windy	0.048
Temp	0.029

First-Level Split on: Outlook as it has the highest information gain.

Decision Tree Construction (Overall steps)

1. **Root Node:**
   - Choose feature with **highest information gain** → `Outlook`
2. **Split on Outlook:**
   - `Sunny` → Subset → [5 samples]
   - `Overcast` → Subset → [4 samples]
   - `Rainy` → Subset → [5 samples]
3. **Leaf Nodes / Further Splits:**
   - `Outlook = Overcast` → All Play = Yes → **Leaf = Yes**
   - `Outlook = Sunny` → Use `Humidity` (best IG in this subset)
   - `Outlook = Rainy` → Use `Windy` (best IG in this subset)

First-Level Split on: Outlook (as outlook gives the highest information gain)

We divide the full dataset into **three subsets** based on values of `Outlook`:

- `Sunny`
- `Overcast`
- `Rainy`

Subset where Outlook = Sunny

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
sunny	mild	high	false	no
sunny	cool	normal	false	yes
sunny	mild	normal	true	yes

Used for second-level split under the Sunny branch.

Subset where Outlook = Overcast

Outlook	Temperature	Humidity	Windy	Play
overcast	hot	high	false	yes
overcast	cool	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes

All Play = Yes → this is a pure leaf node (*no further split needed*).

Subset where Outlook = Rainy

Outlook	Temperature	Humidity	Windy	Play
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
rainy	mild	normal	false	yes
rainy	mild	high	true	no

Used for second-level split under the Rainy branch

Second-Level Splits

Branch: Outlook = Sunny → 5 samples (2 yes, 3 no)

Temperature	Humidity	Windy	Play
hot	high	false	no
hot	high	true	no
mild	high	false	no
cool	normal	false	yes
mild	normal	true	yes

$$Entropy(Sunny) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} \approx 0.971$$

$$Information\ Gain(Sunny, Temperature) = 0.571$$
$$Information\ Gain(Sunny, Humidity) = 0.971$$
$$Information\ Gain(Sunny, Windy) = 0.020$$

Best Attribute for Sunny = **Humidity** (Gain = 0.971)

Calculations shown on next slide

Information gain at second level – Calculations (Outlook = Sunny)

(a) Temperature

Values: `hot`, `mild`, `cool`

- **hot**: 2 samples → `no=2`, `yes=0`
  Entropy = 0
- **mild**: 2 samples → `no=1`, `yes=1`
  Entropy = $-0.5\log_2 0.5 - 0.5\log_2 0.5 = 1.0$
- **cool**: 1 sample → `yes=1`, `no=0`
  Entropy = 0

Weighted entropy:

$$= \frac{2}{5}(0) + \frac{2}{5}(1.0) + \frac{1}{5}(0) = 0.4$$

Information gain:

$$IG(Temperature) = 0.971 - 0.4 = 0.571$$

(b) Humidity

Values: `high`, `normal`

- **high**: 3 samples → `no=3`, `yes=0` → Entropy = 0
- **normal**: 2 samples → `no=0`, `yes=2` → Entropy = 0

Weighted entropy:

$$= \frac{3}{5}(0) + \frac{2}{5}(0) = 0$$

Information gain:

$$IG(Humidity) = 0.971 - 0 = 0.971$$

Information gain at second level – Calculations (Outlook = Sunny)

(c) Windy

Values: `false`, `true`

- **false**: 3 samples → (`no=2`, `yes=1`)
  $Entropy = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \approx 0.918$
- **true**: 2 samples → (`no=1`, `yes=1`)
  $Entropy = 1.0$

Weighted entropy:

$$= \frac{3}{5}(0.918) + \frac{2}{5}(1.0) = 0.5508 + 0.4 = 0.9508$$

Information gain:

$$IG(Windy) = 0.971 - 0.9508 = 0.0202$$

Attribute	IG
Temperature	0.571
Humidity	0.971
Windy	0.020

The **best split** is on **Humidity** (highest IG = 0.971).

Branch: Outlook = Rainy → total: 5 samples (3 yes, 2 no)

Temperature	Humidity	Windy	Play
mild	high	false	yes
cool	normal	false	yes
cool	normal	true	no
mild	normal	false	yes
mild	high	true	no

$$Entropy(Rainy) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} \approx 0.971$$

Information Gain(Sunny, Humidity) = 0.171
Information Gain(Sunny, Temperature) = 0.020
Information Gain(Sunny, Windy) = 0.971

Best Attribute for Sunny = **Windy** (Gain = 0.971)

Calculations shown on next slide

We compute **information gain** again.

- Best split: `Windy`
  - `Windy = False` → 3 Yes
  - `Windy = True` → 2 No

Split on Windy

New branches:

- `Windy = False` → **Play = Yes**
- `Windy = True` → **Play = No**

Both are pure → Stop here.

Information gain at second level – Calculations (Outlook = Rainy)

◆ **Attribute:** ***Humidity***

- **High**: 1 sample (1 yes) → Entropy = 0
- **Normal**: 4 samples (2 yes, 2 no)

$$Entropy(normal) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1.0$$

$$Information\ Gain(Rainy,\ Humidity) = 0.971 - \left(\frac{1}{5}\cdot 0 + \frac{4}{5}\cdot 1.0\right) = 0.971 - 0.8 = \boxed{0.171}$$

◆ **Attribute:** ***Temperature***

- **Mild**: 3 samples (2 yes, 1 no)

$$Entropy(mild) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \approx 0.918$$

- **Cool**: 2 samples (1 yes, 1 no) → Entropy = 1.0

$$Information\ Gain(Rainy,\ Temperature) = 0.971 - \left(\frac{3}{5}\cdot 0.918 + \frac{2}{5}\cdot 1.0\right) = 0.971 - (0.551 + 0.4) = \boxed{0.020}$$

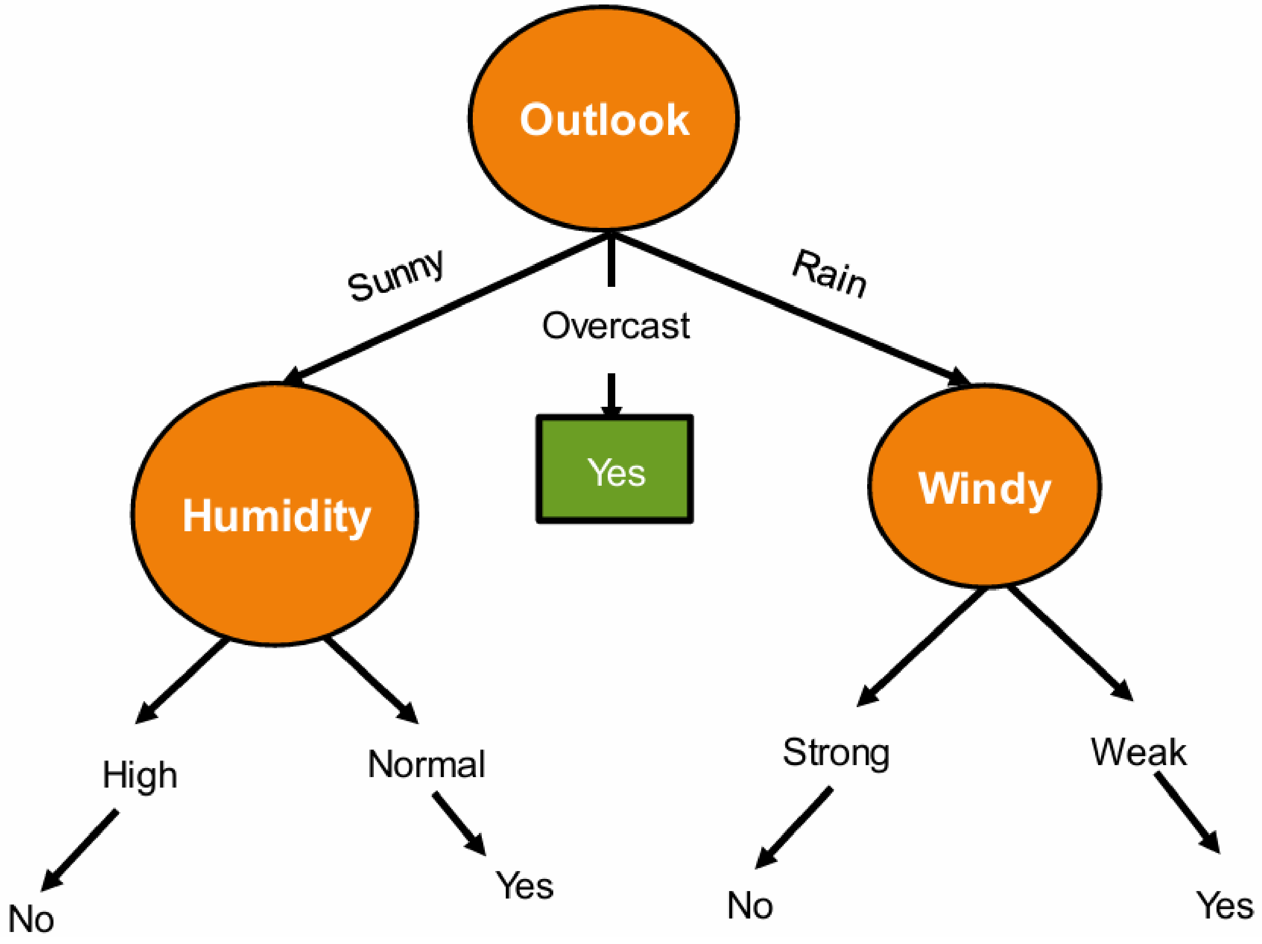
◆ **Attribute:** ***Windy***

- **False**: 3 samples (3 yes) → Entropy = 0
- **True**: 2 samples (0 yes, 2 no) → Entropy = 0

$$Information\ Gain(Rainy,\ Windy) = 0.971 - \left(\frac{3}{5}\cdot 0 + \frac{2}{5}\cdot 0\right) = \boxed{0.971}$$

Final Decision Tree

```
Outlook?
├── Sunny
│    └── Humidity?
│         ├── High → No
│         └── Normal → Yes
├── Overcast → Yes
└── Rainy
     └── Windy?
          ├── False → Yes
          └── True → No
```



End