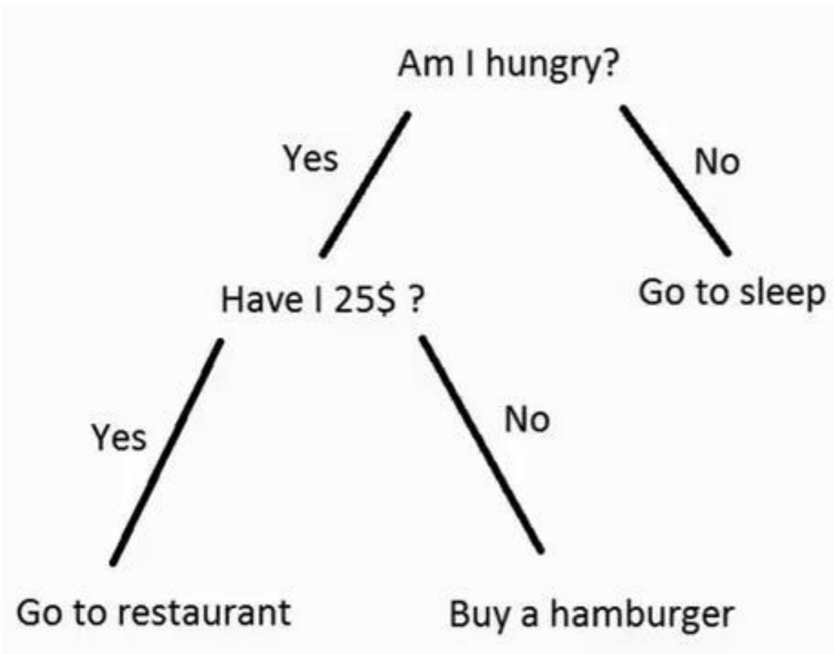
Artificial Intelligence (CS-401)

Decision Trees

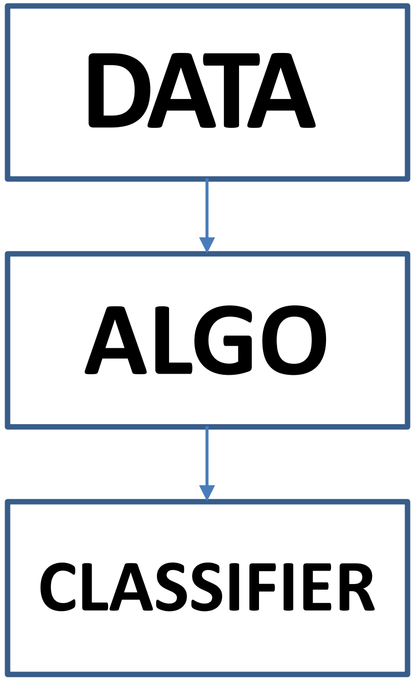
What is a decision tree?

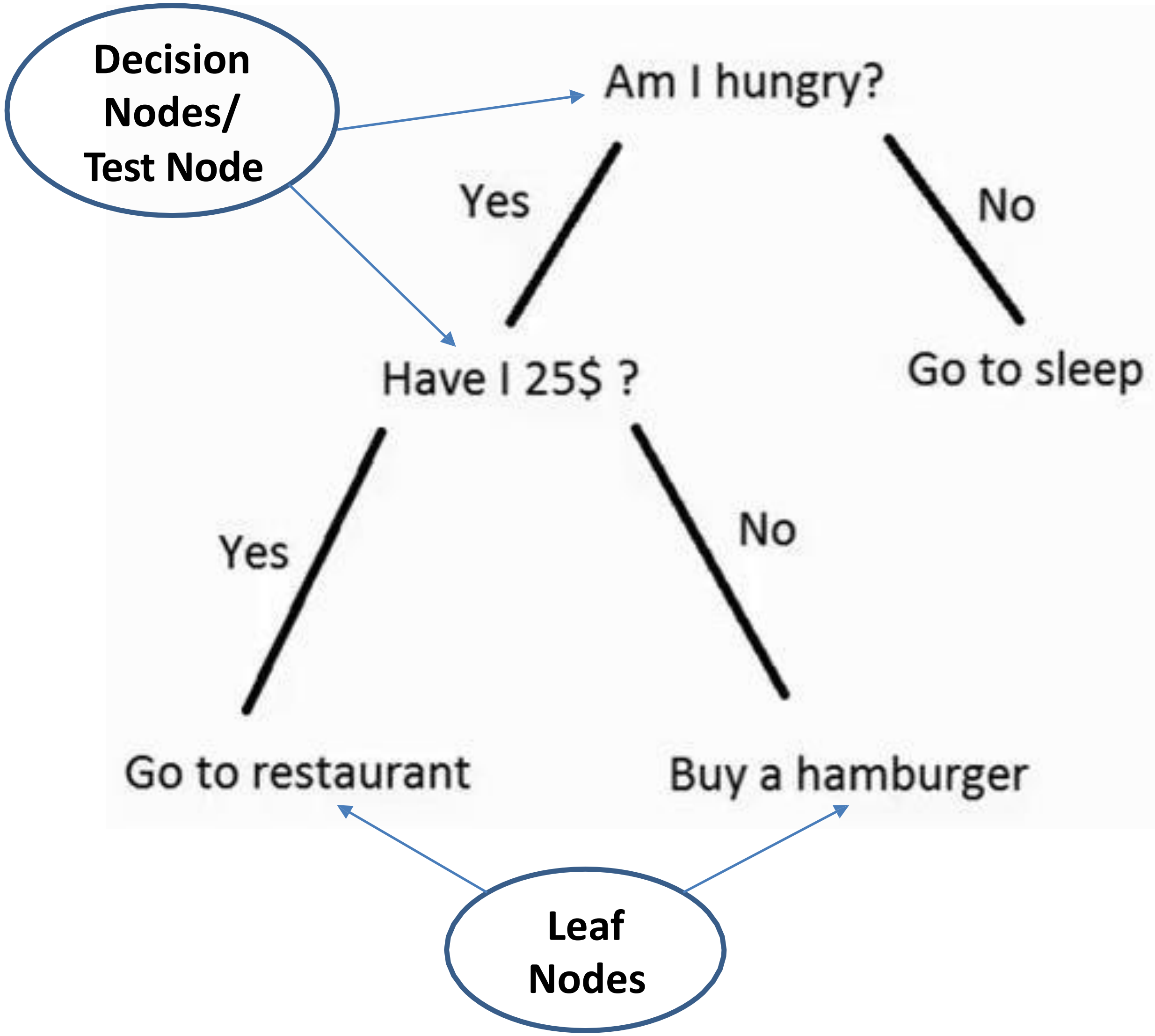
- Decision tree is a graphical representation of all the possible solutions / decisions related to a given problem.
- Decision trees can employed in “classification” and “regression” problems.
- Classification is usually performed using decision trees classifiers.



What is a decision tree?

- It is a tree structures classifier

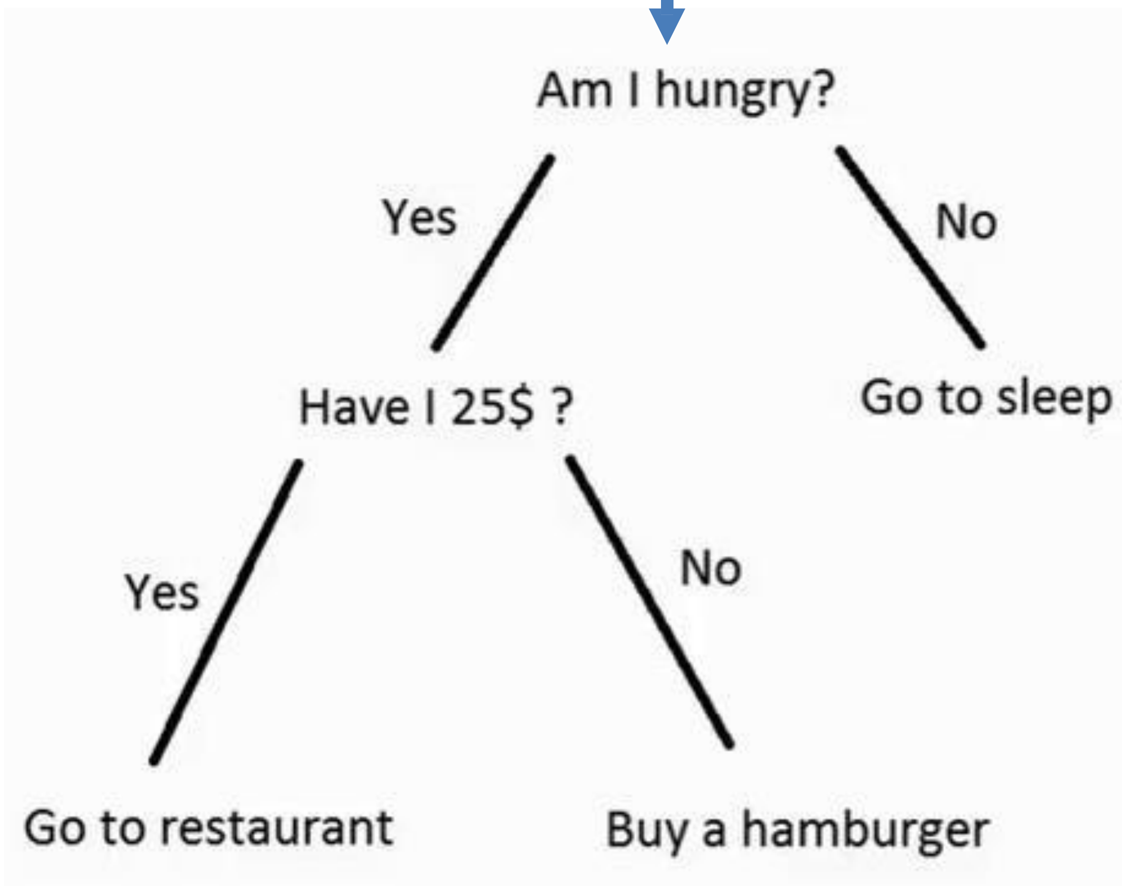




What is a decision tree?

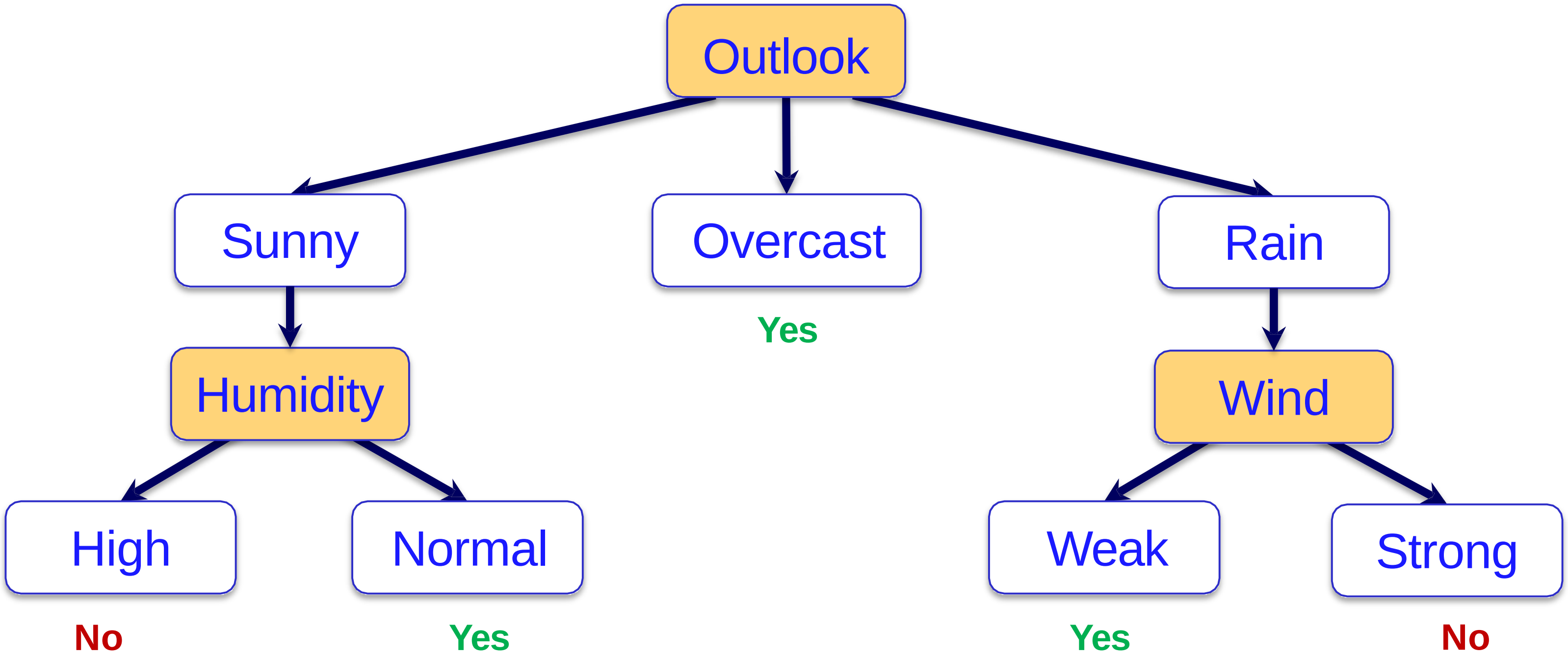
- Data Split
- Decision trees can employed in “classification” and “regression” problems.
- Classification is usually performed using decision trees classifiers.

Am I Hungry	Have I $25	Go to Sleep
Yes		
Yes		
No		
Yes		
No		



Example of a Decision Tree

Is it a good weather to play outside?



How to learn such a tree from past experience?

Decision Tree – Motivation Example

Given the following training examples, will you play in D15?

Divide and conquer:

- split into subsets
- are they pure?
  (all yes or all no)
- if yes: stop
- if not: repeat

See which subset new data falls into

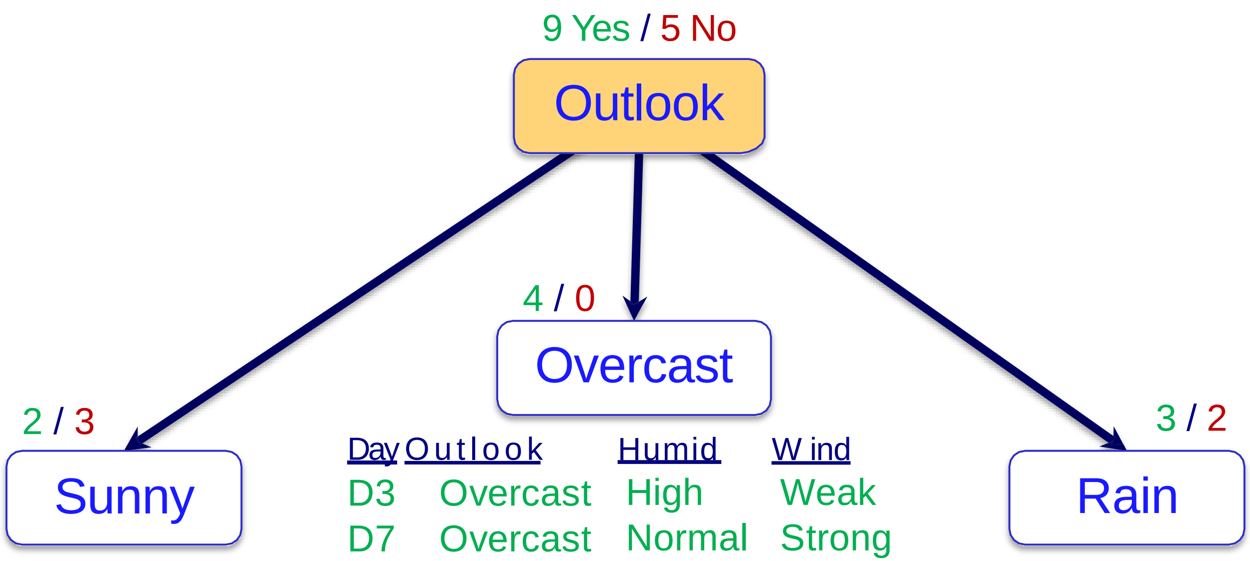
Training Examples

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

New data:

D15	**Rain**	**High**	**Weak**	**???**

Decision Tree – Motivation Example

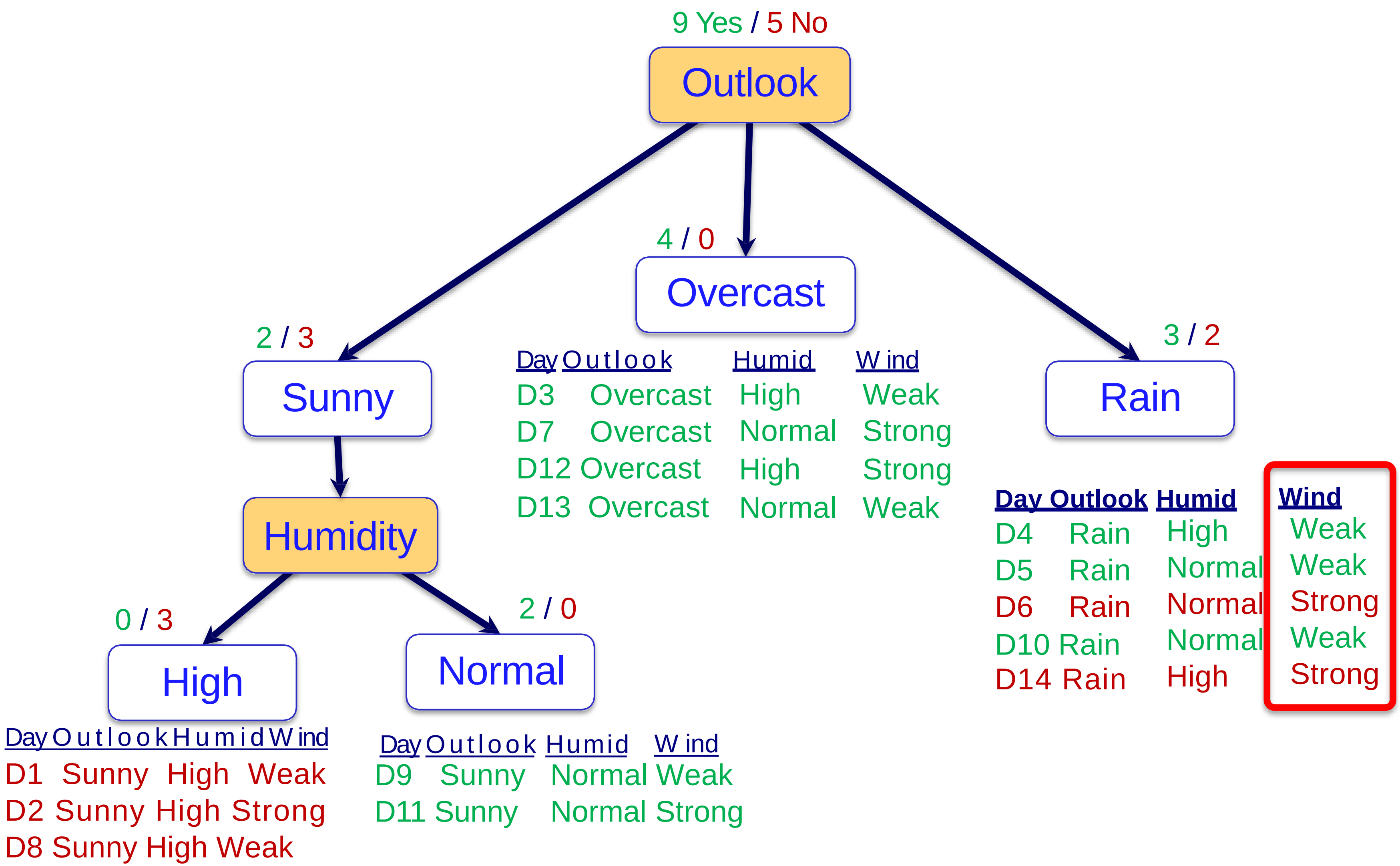


Day	Outlook	Humid	Wind
D1	Sunny	High	Weak
D2	Sunny	High	Strong
D8	Sunny	High	Weak
D9	Sunny	Normal	Weak
D11	Sunny	Normal	Strong

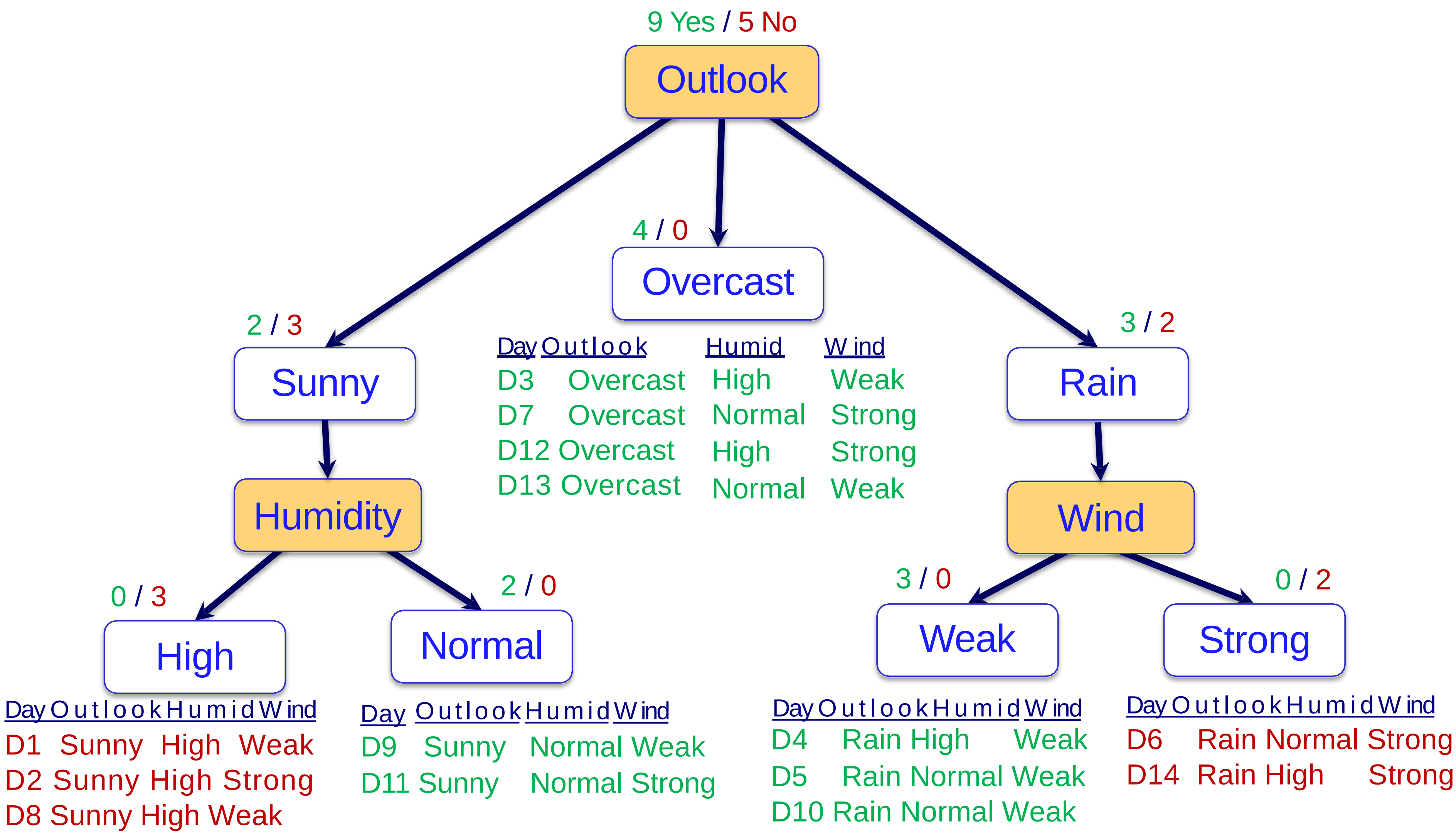
Day	Outlook	Humid	Wind
D3	Overcast	High	Weak
D7	Overcast	Normal	Strong
D12	Overcast	High	Strong
D13	Overcast	Normal	Weak

Day	Outlook	Humid	Wind
D4	Rain	High	Weak
D5	Rain	Normal	Weak
D6	Rain	Normal	Strong
D10	Rain	Normal	Weak
D14	Rain	High	Strong

Decision Tree – Motivation Example

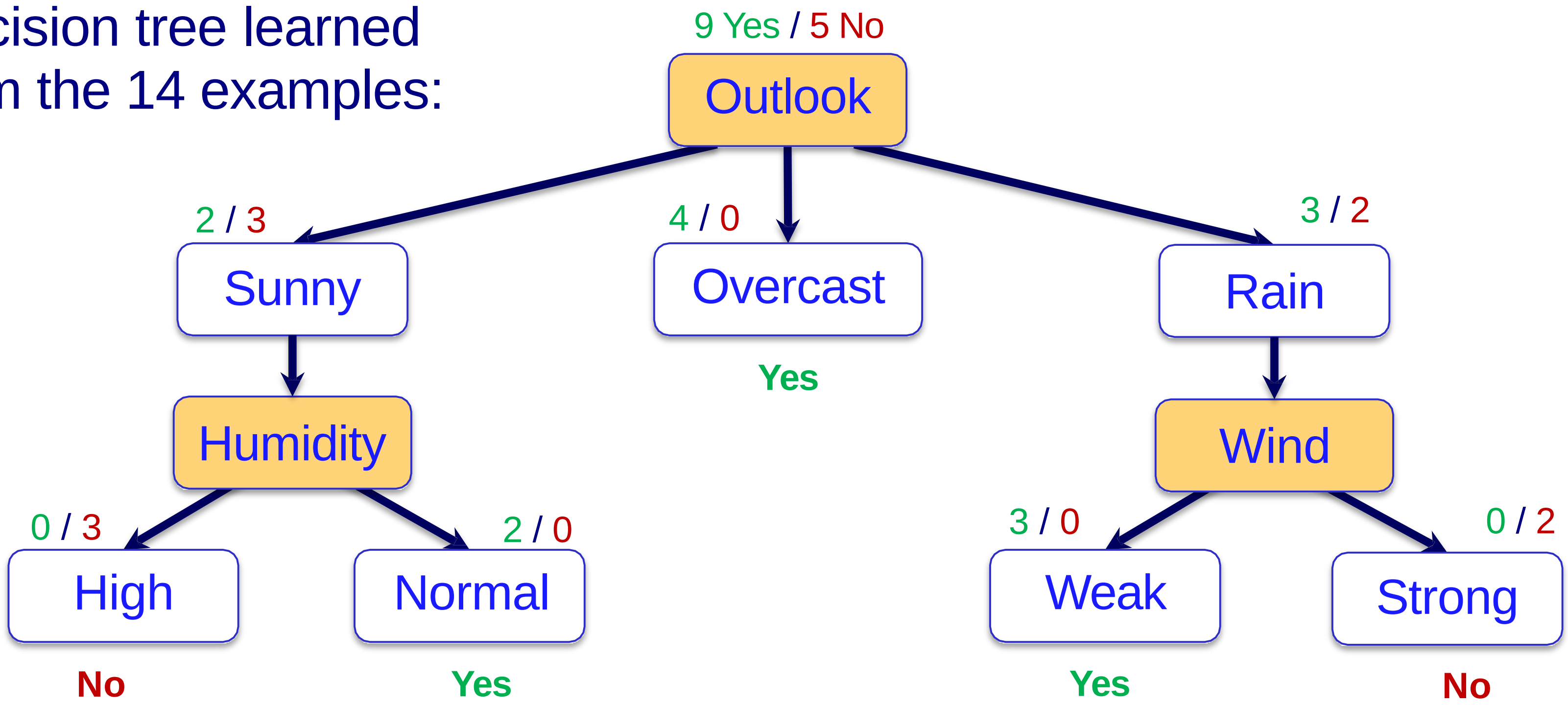


Decision Tree – Motivation Example



Decision Tree – Motivation Example

Decision tree learned from the 14 examples:



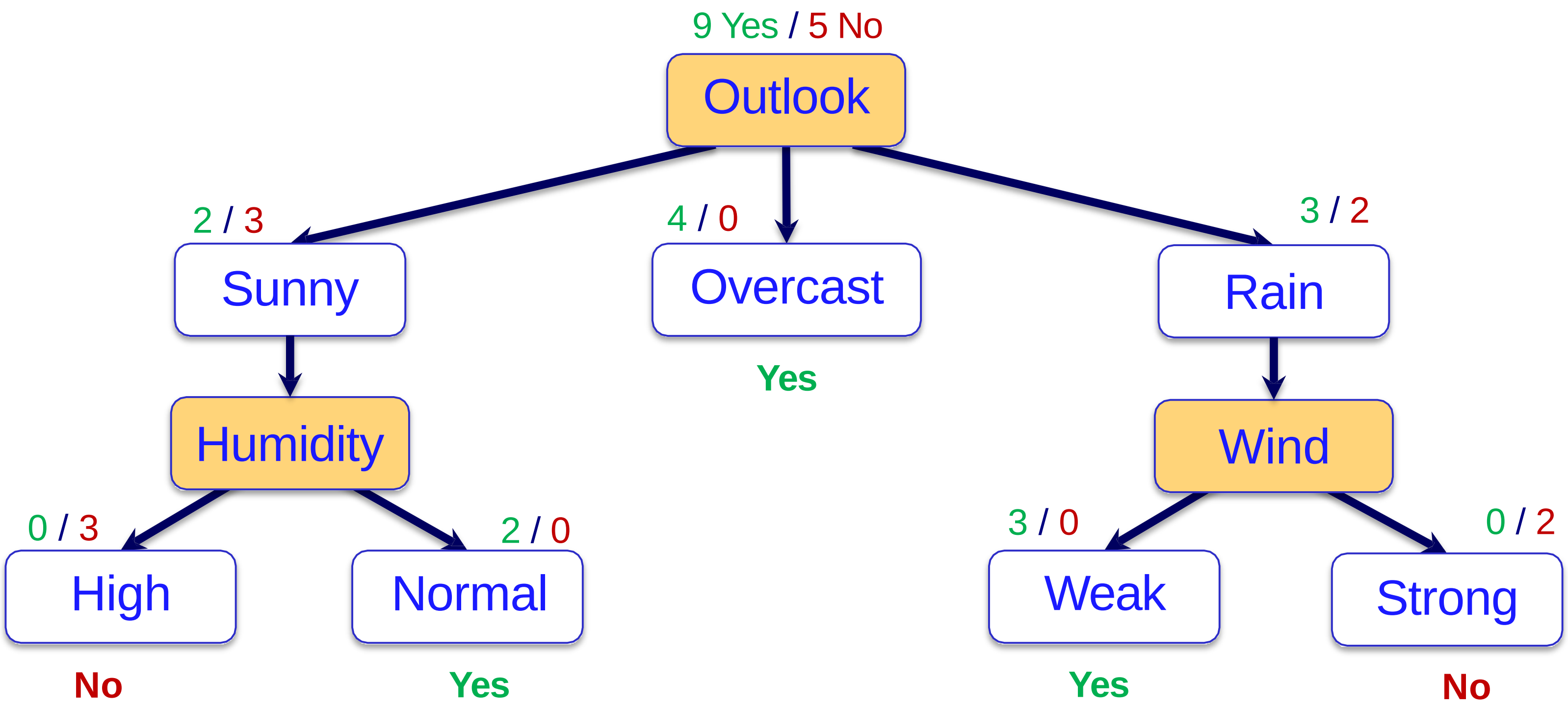
Decision Rule:

$$\text{Yes} \Leftrightarrow (\text{Outlook=Overcast}) \vee (\text{Outlook=Sunny} \wedge \text{Humidity=Normal}) \vee (\text{Outlook=Rain} \wedge \text{Wind=Weak})$$

Day	Outlook	Humid	Wind	
D15	**Rain**	**High**	**Weak**	**???**

➔ Play

Decision Trees are Interpretable

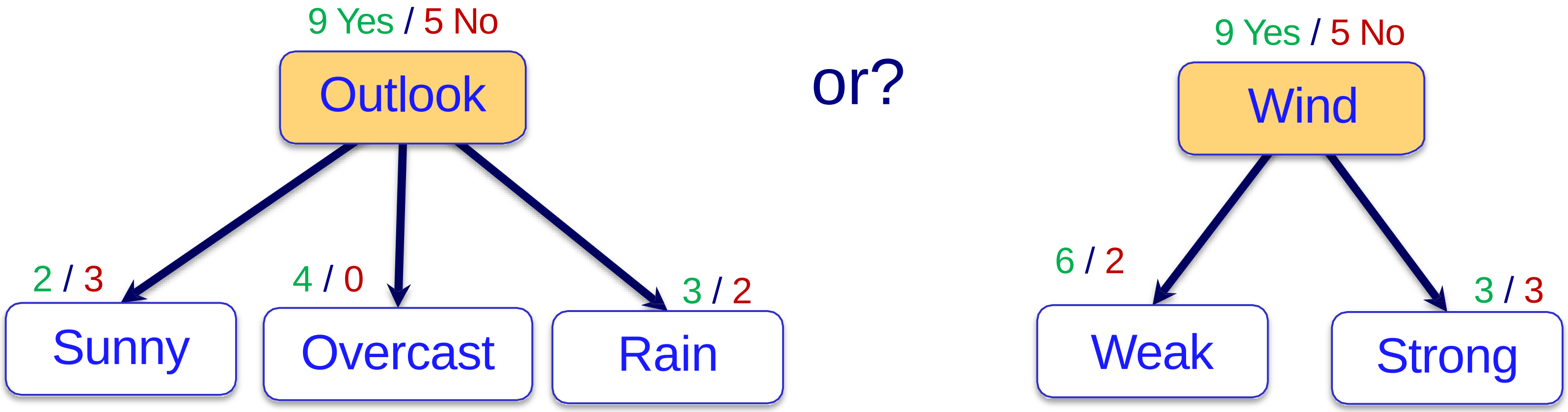


Disjunction of conjunctions of constraints on the attribute values of instances i.e., (... ∧... ∧...) ∨(... ∧... ∧...) ∨...

Set of **if-then-rules**, each branch represents one if-then-rule

- **if-part**: conjunctions of attribute tests on the nodes
- **then-part**: classification of the branch

Which attribute to split on?



Want to measure "purity" of the split

- more certain about Yes/No after the split
  - pure set (4 yes / 0 no) => completely certain (100%)
  - impure (3 yes / 3 no) => completely uncertain (50%)

- can't use the probability of "yes" given the set, P("yes" | set):
  - must be symmetric: 4 yes / 0 no as pure as 0 yes / 4 no

Entropy

Entropy tells us how much a set of data is pure/impure

For binary classification:

$$\text{Entropy}(S) = H(S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus} \text{ bits}$$

– S … is a sample training examples

– $p_{\oplus}$ proportion of positive examples in S

– $p_{\ominus}$ proportion of negative examples in S

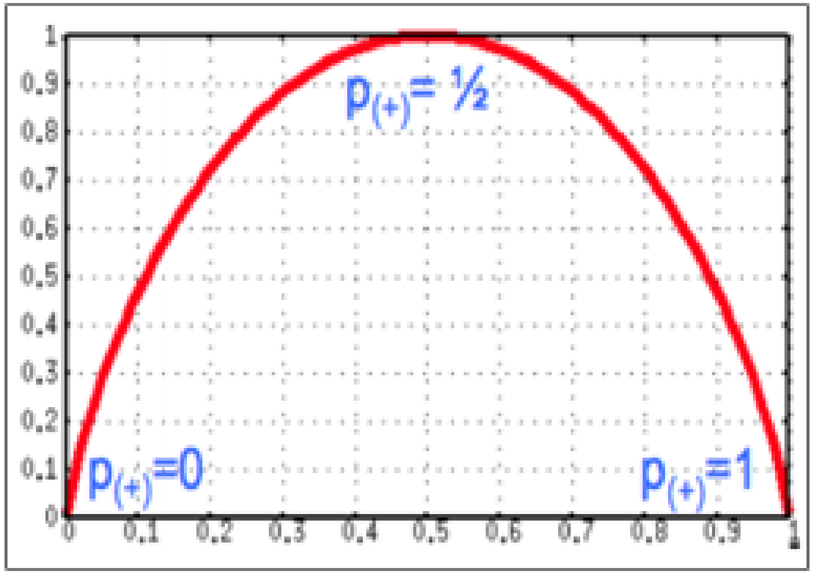
– $p_{\oplus} / p_{\ominus}$ … % of positive/negative examples in S

- impure (3 yes / 3 no):

$$H(S) = -\frac{\%}{\&}\log_2\frac{\%}{\&} - \frac{\%}{\&}\log_2\frac{\%}{\&} = 1 \text{ bits}$$

- pure set (4 yes / 0 no):

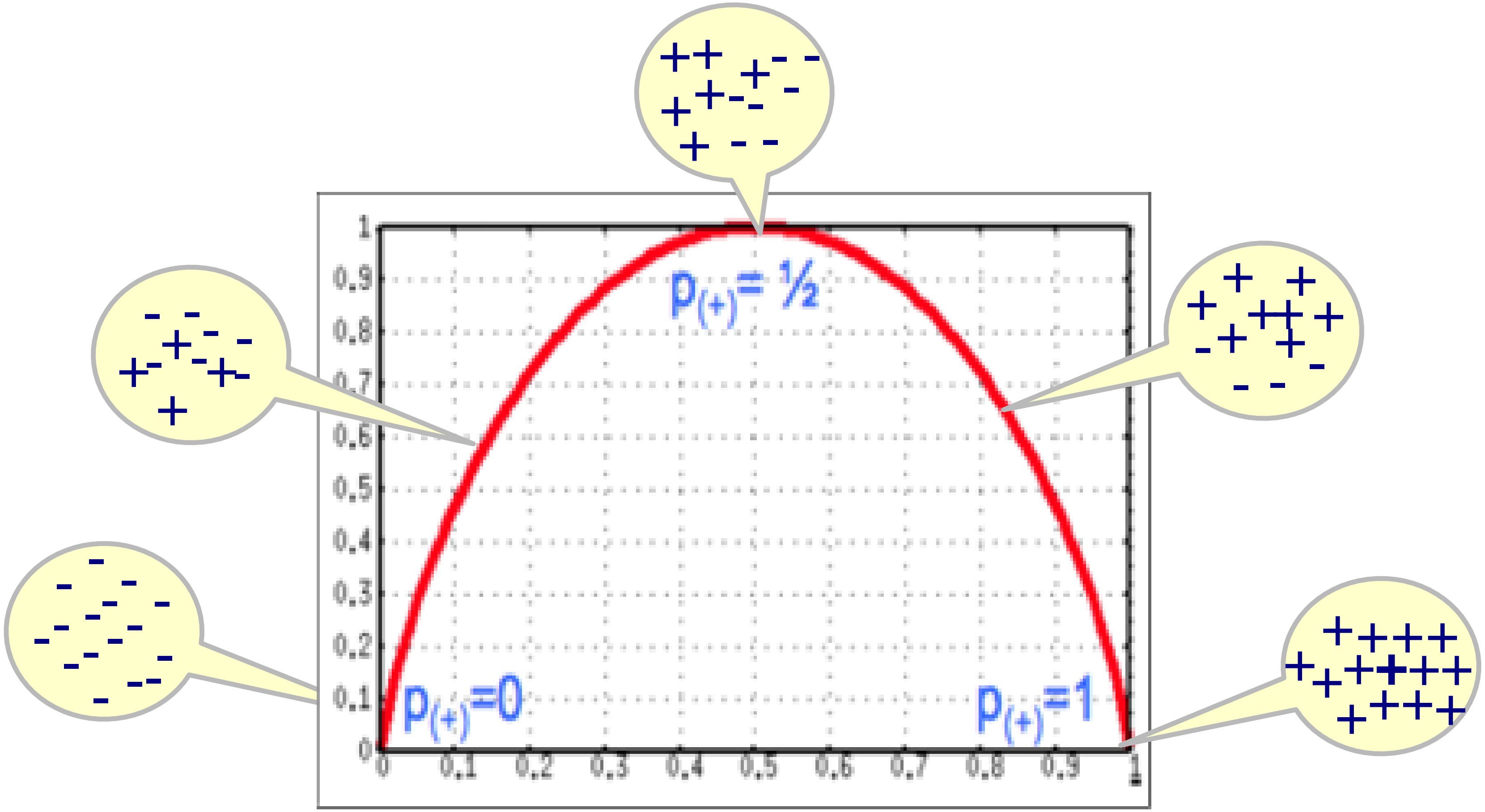
$$H(S) = -\frac{!}{!}\log_2\frac{!}{!} - \frac{"}{!}\log_2\frac{"}{!} = 0 \text{ bits}$$



Entropy

Entropy tells us how much a set of data is pure/impure

$$= -\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$



Information Gain

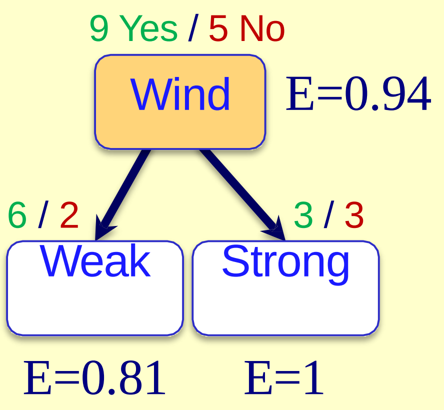
Entropy measures purity at each node, information gain looks at all nodes together and the expected drop in entropy after split.

Gain(S, A) = expected reduction in entropy due to sorting on A

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

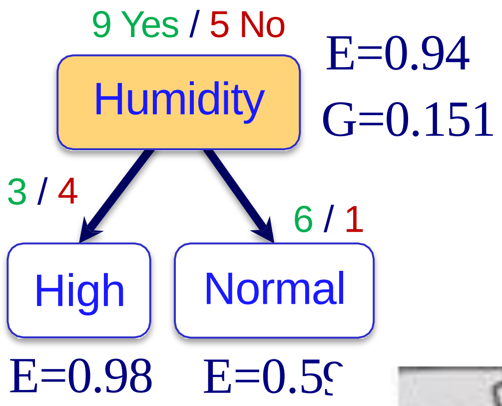
Maximum Gain(S, A) is selected!

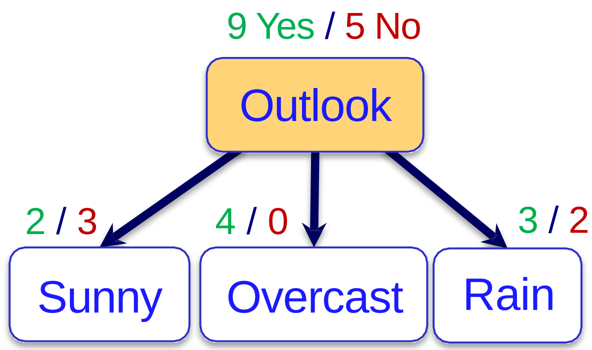
Example

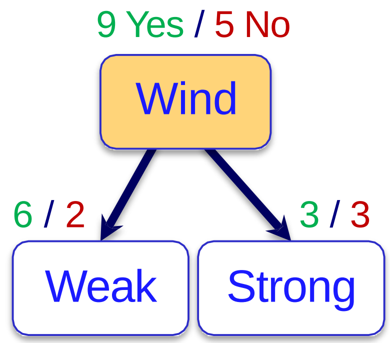


$$Gain\,(\mathrm{S, Wind})$$
$$= 0.94 - (8/14)\cdot 0.81 - (6/14)\cdot 1$$
$$= 0.048$$

Full Example (Entropy & Information Gain)







$$= -\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

Entropy (Humidity) = -9/14·log$_2$(9/14) - 5/14·log$_2$(5/14) = 0.94

Entropy (High) = -3/7·log$_2$(3/7) - 4/7·log$_2$(4/7) = 0.98

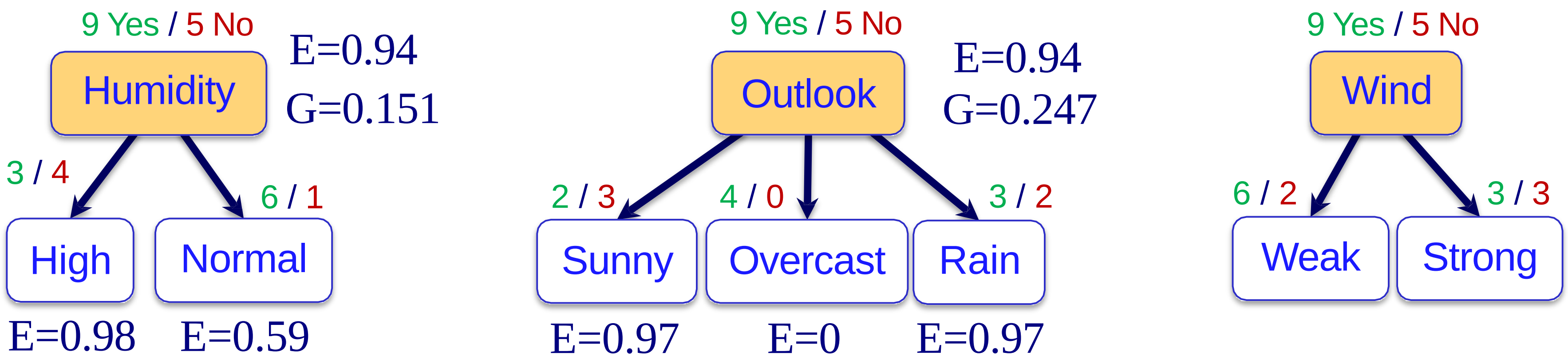
Entropy (Normal) = -6/7·log$_2$(6/7) - 1/7·log$_2$(1/7) = 0.59

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

Gain (S, Humidity) = 0.94 - (7/14)· 0.98 - (7/14) · 0.59 = 0.151

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

Full Example (Entropy & Information Gain)



$$Entropy\ (\text{Outlook}) = -9/14 \cdot \log_2(9/14) - 5/14 \cdot \log_2(5/14) = 0.94$$

$$Entropy(\text{Sunny}) = -2/5 \cdot \log_2(2/5) - 3/5 \cdot \log_2(3/5) = 0.97$$

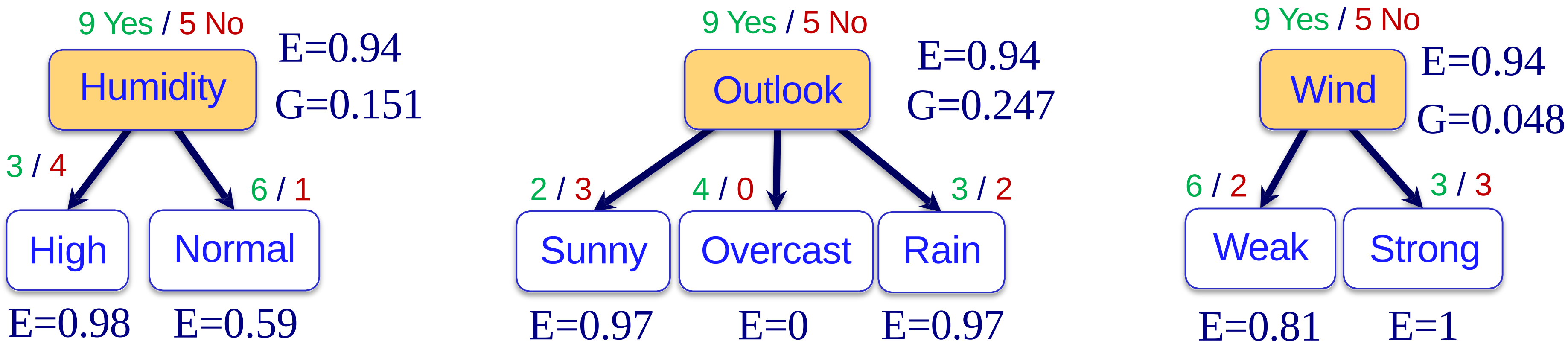
$$Entropy(\text{Overcast}) = -4/4 \cdot \log_2(4/4) - 0/4 \cdot \log_2(0/4) = 0$$

$$Entropy(\text{Rain}) = -3/5 \cdot \log_2(3/5) - 2/5 \cdot \log_2(2/5) = 0.97$$

$$Gain\ (S, \text{Outlook}) = 0.94 - (5/14) \cdot 0.97 - (4/14) \cdot 0 - (5/14) \cdot 97 = 0.247$$

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

Full Example (Entropy & Information Gain)



$$\textit{Entropy}\ (\text{Wind}) = -9/14 \cdot \log_2(9/14) - 5/14 \cdot \log_2(5/14) = 0.94$$

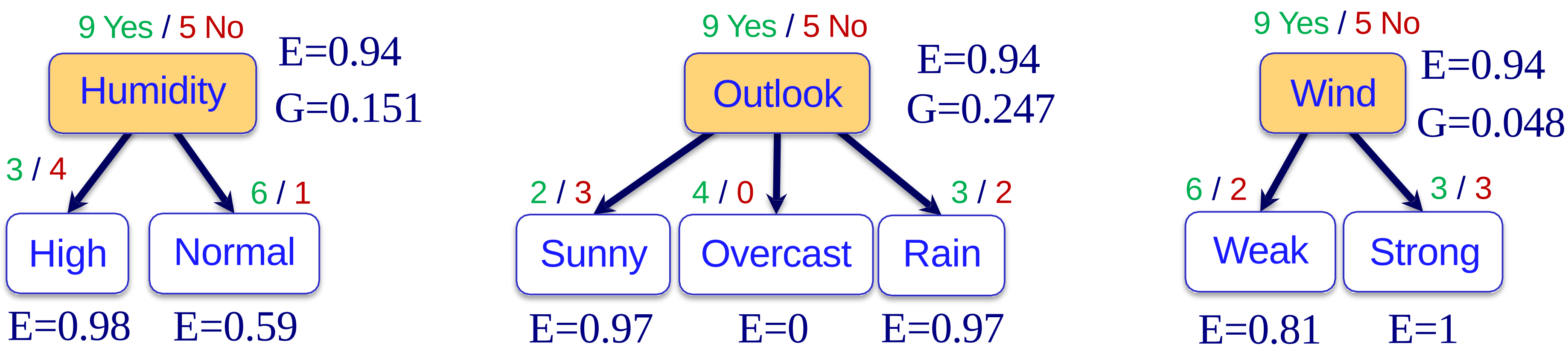
$$\textit{Entropy}\ (\text{Weak}) = -6/8 \cdot \log_2(6/8) - 2/8 \cdot \log_2(2/8) = 0.81$$

$$\textit{Entropy}\ (\text{Strong}) = -3/6 \cdot \log_2(3/6) - 3/6 \cdot \log_2(3/6) = 1$$

$$\textit{Gain}\ (S, \text{Wind}) = 0.94 \quad - (8/14) \cdot 0.81 - (6/14) \cdot 1 = 0.048$$

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

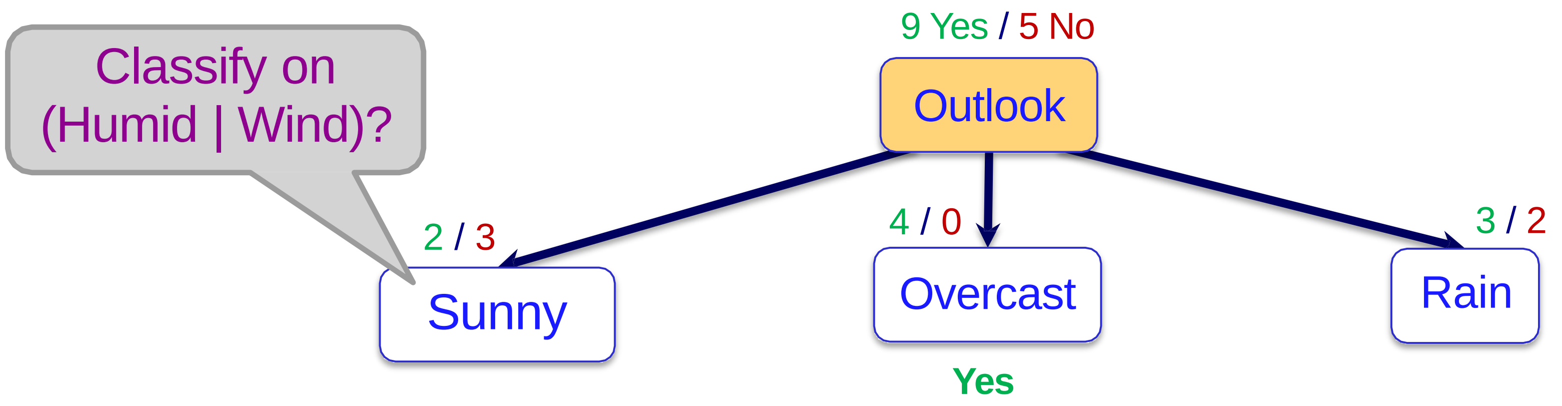
Full Example (Entropy & Information Gain)

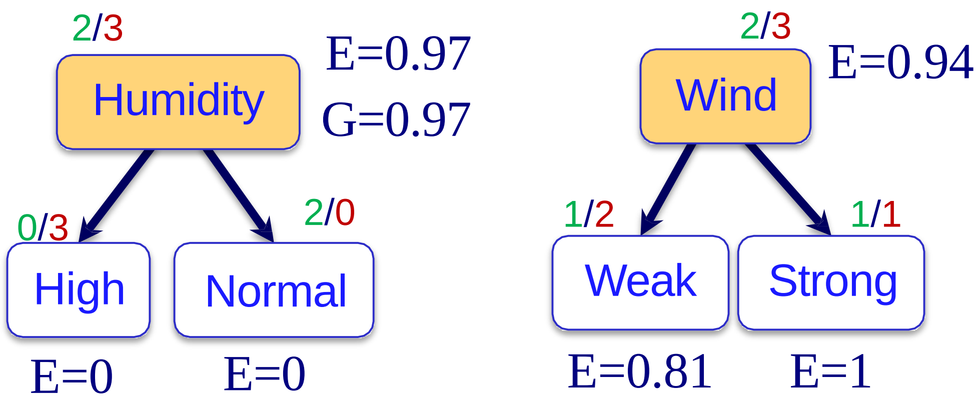


The attribute with the largest Information Gain (Outlook 0.247) is selected as the decision node.

Nodes with zero Entropy (e.g., Overcast) does not need splitting

Full Example (Entropy & Information Gain)





Day	Outlook	Humid	Wind
D1	Sunny	High	Weak
D2	Sunny	High	Strong
D8	Sunny	High	Weak
D9	Sunny	Normal	Weak
D11	Sunny	Normal	Strong

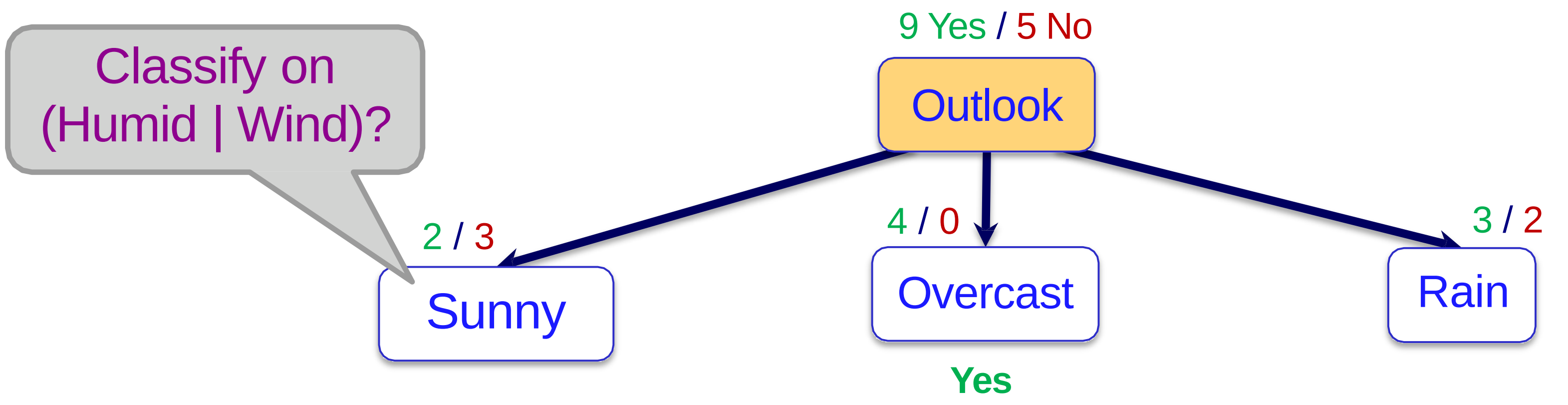
$$\textit{Entropy}\ (\text{Humidity}) = -2/5 \cdot \log_2(2/5) - 3/5 \cdot \log_2(3/5) = 0.97$$

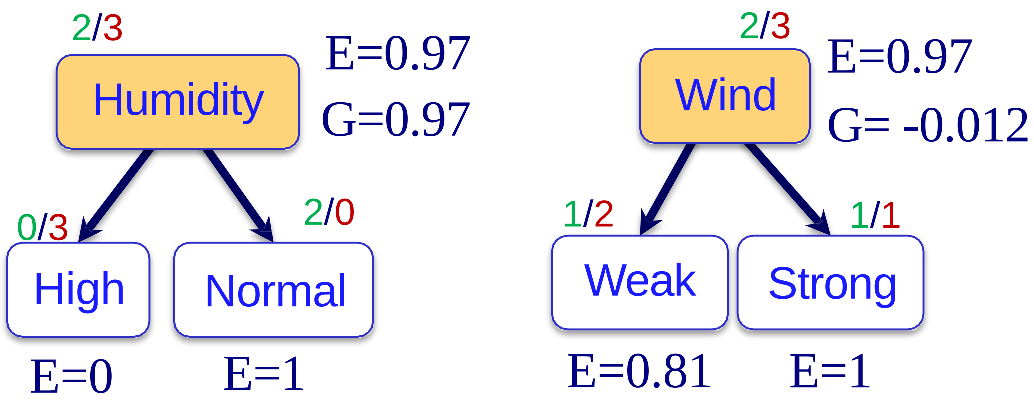
$$\textit{Entropy}\ (\text{High}) = -0/3 \cdot \log_2(0/3) - 3/3 \cdot \log_2(3/3) = 0$$

$$\textit{Entropy}\ (\text{Normal}) = -2/2 \cdot \log_2(2/2) - 0/2 \cdot \log_2(0/2) = 0$$

$$\textit{Gain}\ (\text{S, Humidity}) = 0.97 - (3/5) \cdot 0 - (2/5) \cdot 0 = 0.97$$

Full Example (Entropy & Information Gain)





Day	Outlook	Humid	Wind
D1	Sunny	High	Weak
D2	Sunny	High	Strong
D8	Sunny	High	Weak
D9	Sunny	Normal	Weak
D11	Sunny	Normal	Strong

$$\textit{Entropy}\ (\text{Wind}) = -2/5 \cdot \log_2(2/5) - 3/5 \cdot \log_2(3/5) = 0.97$$

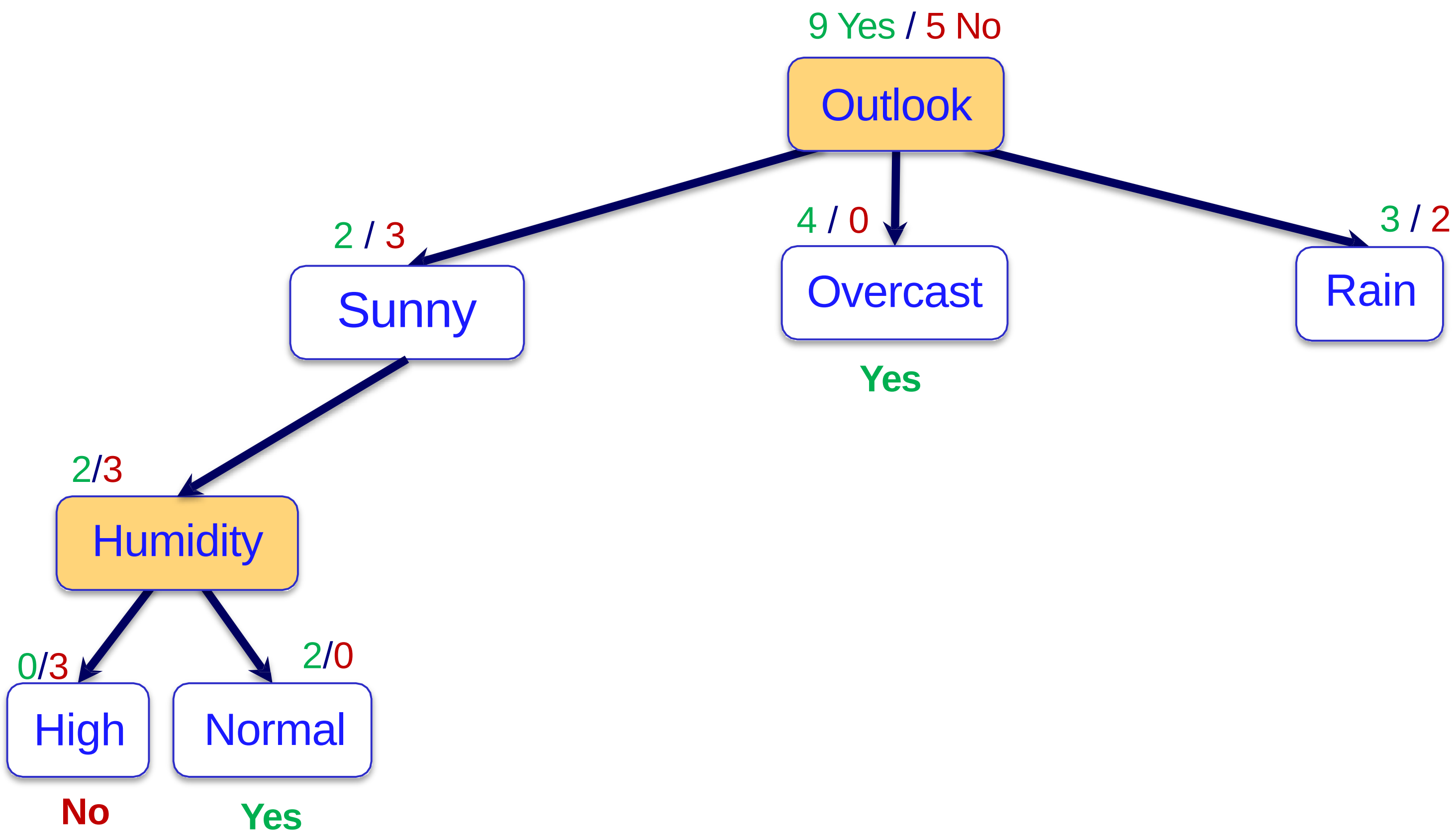
$$\textit{Entropy}\ (\text{Weak}) = -1/3 \cdot \log_2(1/3) - 2/3 \cdot \log_2(2/3) = 0.92$$

$$\textit{Entropy}\ (\text{Strong}) = -1/2 \cdot \log_2(1/2) - 1/2 \cdot \log_2(1/2) = 1$$

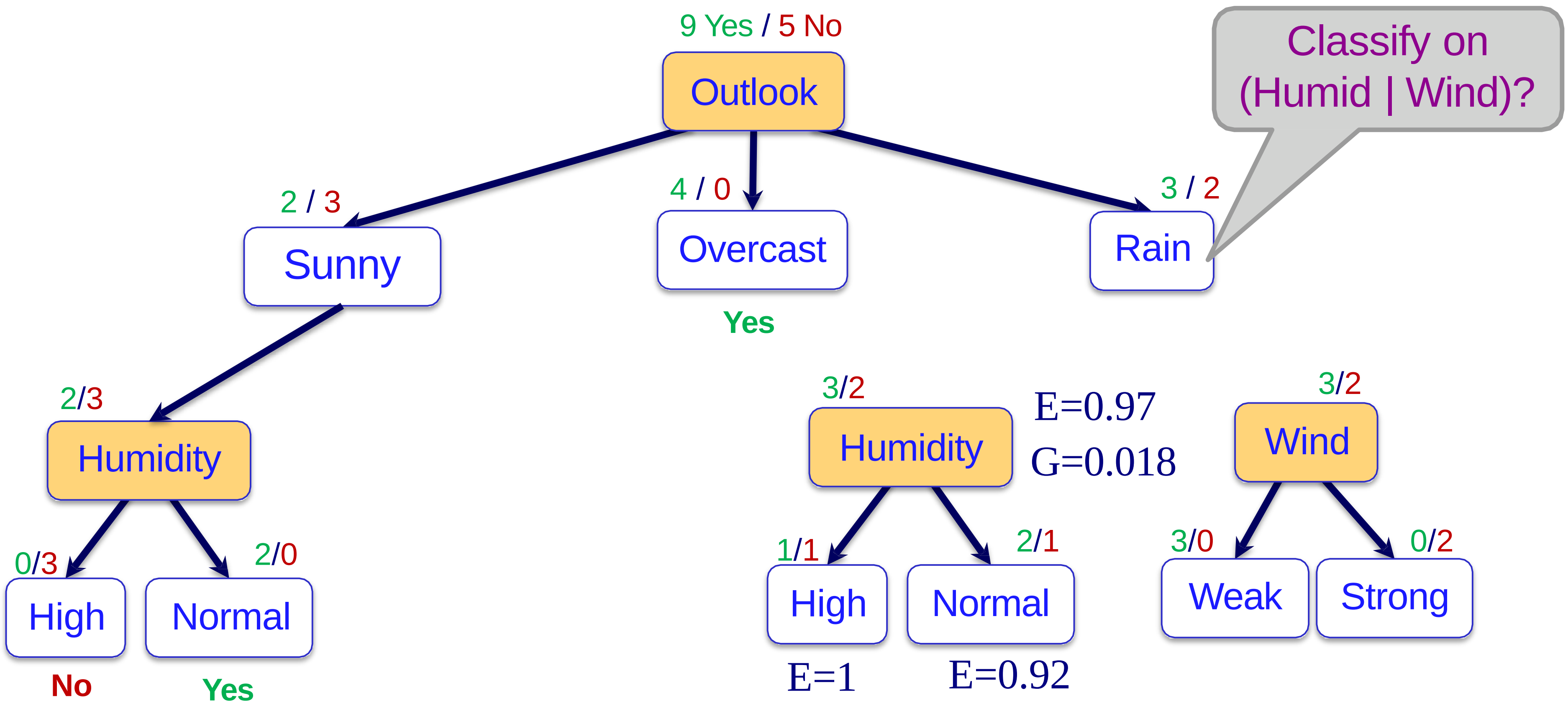
$$\textit{Gain}\ (\text{S, Wind}) = 0.97 \quad - (3/5) \cdot 0.97 - (2/5) \cdot 1 = -0.012$$

➔ Humidity has the highest gain (0.97)

Full Example (Entropy & Information Gain)



Full Example (Entropy & Information Gain)



$$Entropy(\text{Humidity}) = -3/5 \cdot \log_2(3/5) - 2/5 \cdot \log_2(2/5) = 0.97$$

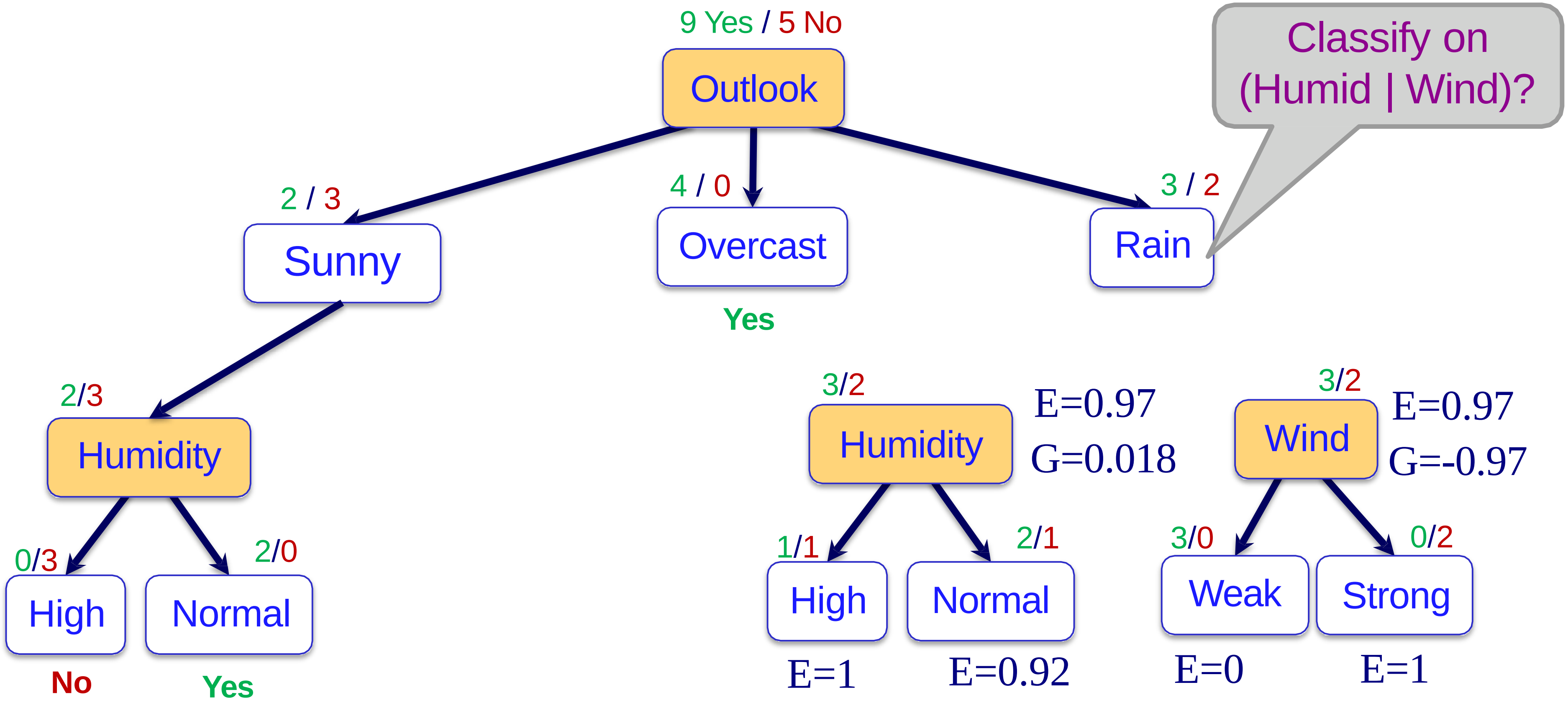
$$Entropy(\text{High}) = -1/2 \cdot \log_2(1/2) - 1/2 \cdot \log_2(1/2) = 1$$

$$Entropy(\text{Normal}) = -2/3 \cdot \log_2(2/3) - 1/3 \cdot \log_2(1/3) = 0.92$$

$$Gain(\text{S, Humidity}) = 0.97 \quad - (2/5) \cdot 1 - (3/5) \cdot 0.92 = 0.018$$

Day	Outlook	Humid	Wind
D4	Rain	High	Weak
D5	Rain	Normal	Weak
D6	Rain	Normal	Strong
D10	Rain	Normal	Weak
D14	Rain	High	Strong

Full Example (Entropy & Information Gain)



$$Entropy(\text{Wind}) = -3/5 \cdot \log_2(3/5) - 2/5 \cdot \log_2(2/5) = 0.97$$

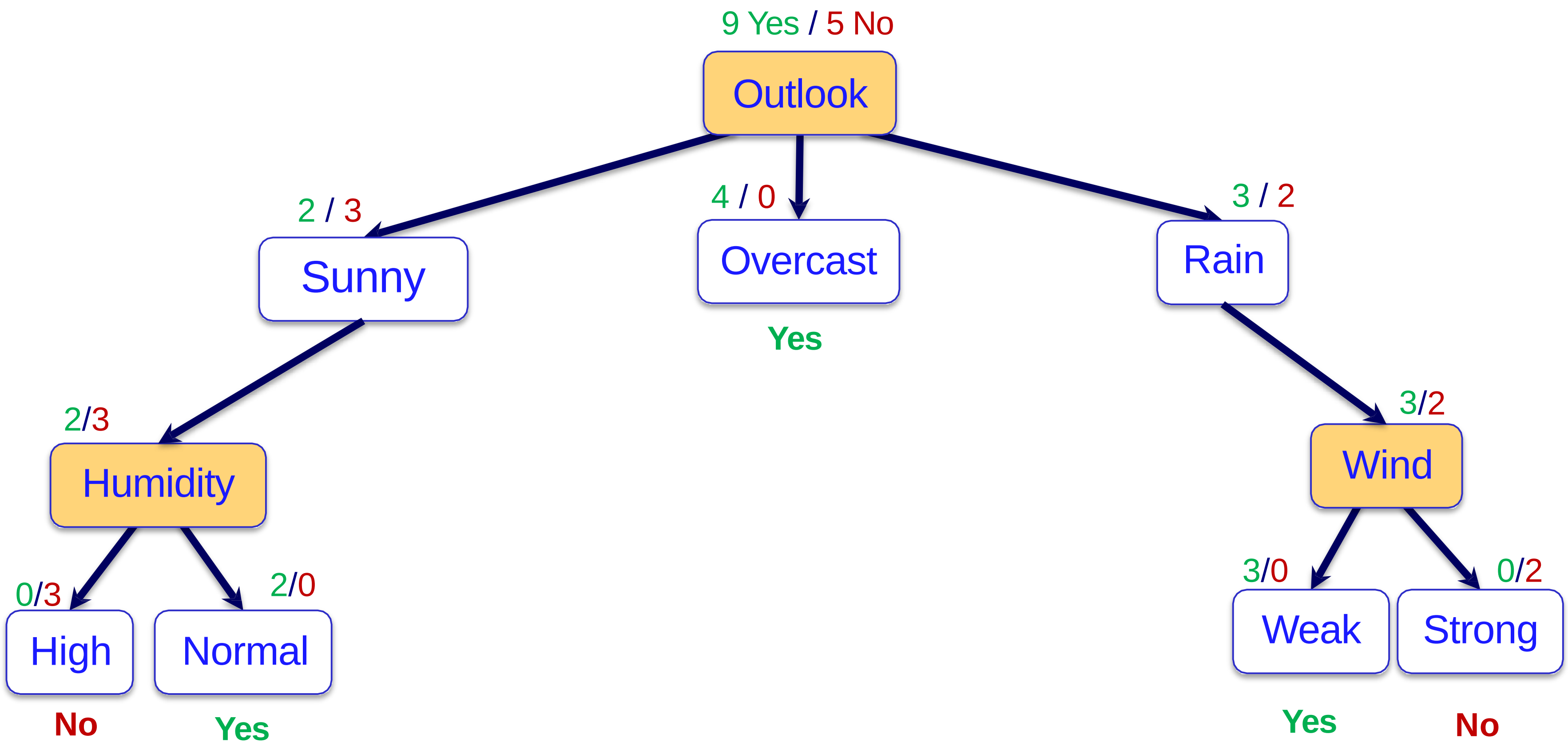
$$Entropy(\text{Weak}) = -3/3 \cdot \log_2(3/3) - 0/3 \cdot \log_2(0/3) = 0$$

$$Entropy(\text{Strong}) = -0/2 \cdot \log_2(0/2) - 2/2 \cdot \log_2(2/2) = 0$$

$$Gain(S, \text{Wind}) = 0.97 \quad - (3/5) \cdot 0 - (2/5) \cdot 0 = 0.97$$

Day	Outlook	Humid	Wind
D4	Rain	High	Weak
D5	Rain	Normal	Weak
D6	Rain	Normal	Strong
D10	Rain	Normal	Weak
D14	Rain	High	Strong

Full Example (Entropy & Information Gain)



Decision Rule:

$$\text{Yes} \Leftrightarrow (\text{Outlook=Overcast}) \vee (\text{Outlook=Sunny} \wedge \text{Humidity=Normal}) \vee (\text{Outlook=Rain} \wedge \text{Wind=Weak})$$

What is good about Decision Trees

- Interpretable: humans can understand decisions
- Easily handles irrelevant attributes (G=0)
- Very compact: # nodes << D after pruning
- Very fast at testing time: O(Depth)

Limitations for Decision Trees

- Greedy (may not find best tree).
- Instances are represented by attribute/value pairs(e.g., Outlook: sunny, Wind: strong), but what if we have discrete input values.
- The target function has discrete output values (e.g., Yes, No), thus we cannot have continues number output values.
- The training data may contain errors, or missing attributes
- Uncertainty in the data (e.g., suppose we have two exact days/features, one with “yes” and one with “no”. ➔ no classifier can help in such totally Uncertain data.