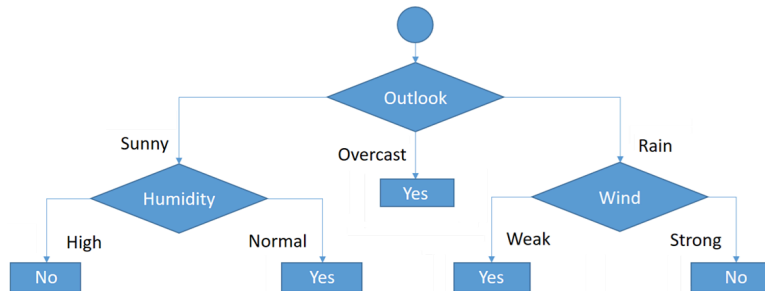


CART Decision Tree Example

Day	Outlook	Temp.	Humidity	Wind	Class
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

The final decision tree is built using Gini Index as shown here.



Outlook	Yes	No	Total
Sunny	2	3	5
Overcast	4	0	4
Rain	3	2	5
Total	9	5	14

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

$$\text{Gini}(\text{Outlook}=\text{Overcast}) = 1 - (4/4)^2 - (0/4)^2 = 0$$

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

Now, calculate weighted sum of gini indexes for outlook feature.

$$\text{Gini}(\text{Outlook}) = (5/14) \times 0.48 + (4/14) \times 0 + (5/14) \times 0.48 = 0.171 + 0 + 0.171 = 0.342$$

Temp.	Yes	No	Total
Hot	2	2	4
Mild	4	2	6
Cool	3	1	4
Total	9	5	14

$$\text{Gini}(\text{Temp}=\text{Hot}) = 1 - (2/4)^2 - (2/4)^2 = 0.5$$

$$\text{Gini}(\text{Temp}=\text{Cool}) = 1 - (3/4)^2 - (1/4)^2 = 1 - 0.5625 - 0.0625 = 0.375$$

$$\text{Gini}(\text{Temp}=\text{Mild}) = 1 - (4/6)^2 - (2/6)^2 = 1 - 0.444 - 0.111 = 0.445$$

Now, calculate weighted sum of gini index for temperature feature

$$\text{Gini}(\text{Temp}) = (4/14) \times 0.5 + (4/14) \times 0.375 + (6/14) \times 0.445 = 0.142 + 0.107 + 0.190 = 0.439$$

Humidity	Yes	No	Total
High	3	4	7
Normal	6	1	7
Total	9	5	14

$$\text{Gini}(\text{Humidity}=\text{High}) = 1 - (3/7)^2 - (4/7)^2 = 1 - 0.183 - 0.326 = 0.489$$

$$\text{Gini}(\text{Humidity}=\text{Normal}) = 1 - (6/7)^2 - (1/7)^2 = 1 - 0.734 - 0.02 = 0.244$$

Weighted sum for humidity feature is calculated as:

$$\text{Gini}(\text{Humidity}) = (7/14) \times 0.489 + (7/14) \times 0.244 = 0.367$$

Wind	Yes	No	Total
Weak	6	2	8
Strong	3	3	6
Total	9	5	14

$$\text{Gini}(\text{Wind}=\text{Weak}) = 1 - (6/8)^2 - (2/8)^2 = 1 - 0.5625 - 0.0625 = 0.375$$

$$\text{Gini}(\text{Wind}=\text{Strong}) = 1 - (3/6)^2 - (3/6)^2 = 1 - 0.25 - 0.25 = 0.5$$

Weighted sum for humidity feature

$$\text{Gini}(\text{Wind}) = (8/14) \times 0.375 + (6/14) \times 0.5 = 0.428$$

Decision for outlook: The Outlook becomes winner (root node) as its Gini index is the lowest.

Feature	Outlook	Temp.	Humidity	Wind
Gini index	0.342	0.439	0.367	0.428

Now, sub-dataset for sunny outlook,

Find Gini index scores for temperature, humidity and wind features respectively.

Gini of temprature for sunny outlook

Temp	Yes	No	Total
Hot	0	2	2
Cool	1	0	1
Mild	1	1	2

$$\text{Gini(Outlook=Sunny and Temp.=Hot)} = 1 - (0/2)^2 - (2/2)^2 = 0$$

$$\text{Gini(Outlook=Sunny and Temp.=Cool)} = 1 - (1/1)^2 - (0/1)^2 = 0$$

$$\text{Gini(Outlook=Sunny and Temp.=Mild)} = 1 - (1/2)^2 - (1/2)^2 = 1 - 0.25 - 0.25 = 0.5$$

$$\text{Gini(Outlook=Sunny and Temp.)} = (2/5) \times 0 + (1/5) \times 0 + (2/5) \times 0.5 = 0.2$$

Gini of humidity for sunny outlook

Humidity	Yes	No	Total
High	0	3	3
Normal	2	0	2

$$\text{Gini(Outlook=Sunny and Humidity=High)} = 1 - (0/3)^2 - (3/3)^2 = 0$$

$$\text{Gini(Outlook=Sunny and Humidity=Normal)} = 1 - (2/2)^2 - (0/2)^2 = 0$$

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

Gini of wind for sunny outlook

Wind	Yes	No	Total
Weak	1	2	3
Strong	1	1	2

$$\text{Gini(Outlook=Sunny and Wind=Weak)} = 1 - (1/3)^2 - (2/3)^2 = 0.266$$

$$\text{Gini(Outlook=Sunny and Wind=Strong)} = 1 - (1/2)^2 - (1/2)^2 = 0.2$$

$$\text{Gini(Outlook=Sunny and Wind)} = (3/5) \times 0.266 + (2/5) \times 0.2 = 0.466$$

Decision for sunny outlook: The winner (root node) is humidity because it has the lowest value.

Feature	Temp.	Humidity	Wind
Gini index	0.2	0	0.466

The decision is always no for high humidity and sunny outlook.

The decision will always be yes for normal humidity and sunny outlook.

Gini of temprature for rain outlook

Temp.	Yes	No	Total
Cool	1	1	2
Mild	2	1	3

$$\text{Gini(Outlook=Rain and Temp.=Cool)} = 1 - (1/2)^2 - (1/2)^2 = 0.5$$

$$\text{Gini(Outlook=Rain and Temp.=Mild)} = 1 - (2/3)^2 - (1/3)^2 = 0.444$$

$$\text{Gini(Outlook=Rain and Temp.)} = (2/5) \times 0.5 + (3/5) \times 0.444 = 0.466$$

Gini of humidity for rain outlook

Humidity	Yes	No	Total
High	1	1	2
Normal	2	1	3

$$\text{Gini(Outlook=Rain and Humidity=High)} = 1 - (1/2)^2 - (1/2)^2 = 0.5$$

$$\text{Gini(Outlook=Rain and Humidity=Normal)} = 1 - (2/3)^2 - (1/3)^2 = 0.444$$

$$\text{Gini(Outlook=Rain and Humidity)} = (2/5) \times 0.5 + (3/5) \times 0.444 = 0.466$$

Gini of wind for rain outlook

Wind	Yes	No	Total
Weak	3	0	3
Strong	0	2	2

$$\text{Gini(Outlook=Rain and Wind=Weak)} = 1 - (3/3)^2 - (0/3)^2 = 0$$

$$\text{Gini(Outlook=Rain and Wind=Strong)} = 1 - (0/2)^2 - (2/2)^2 = 0$$

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

Decision for rain outlook: The winner (root node) is wind feature for rain outlook as it has the minimum gini index score.

Feature	Temp.	Humidity	Wind
Gini index	0.466	0.466	0

The decision is always yes when wind is weak.

The decision is always no if wind is strong.

The final decision tree is built using Gini Index as shown here.