Decision Tree
In Machine Learning

Present By Ahmed Abdo

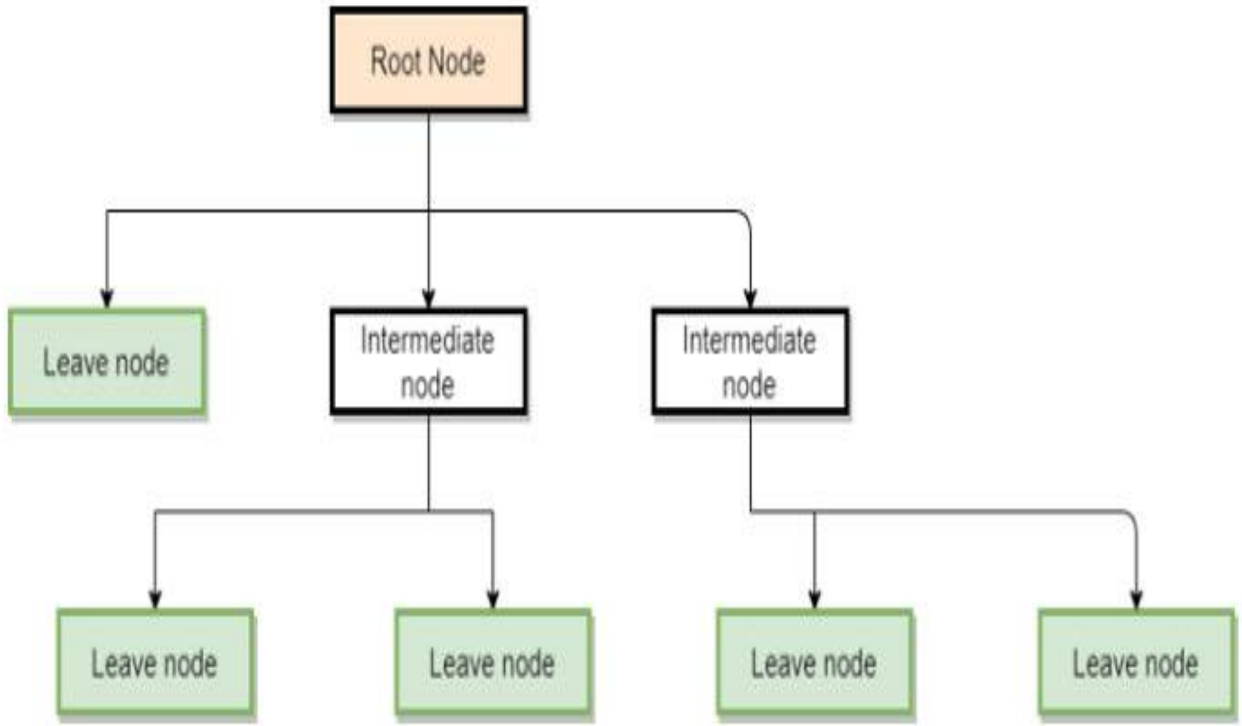
Agenda

- Why do use we Decision Tree?
- How Decision Trees Look like?
- How Decision Trees make a decision?
- How to Determine the Best Split to DT?
- Issues In Decision Tree
- Pruning
- Lab

Why do use we Decision Tree?

- Be a kind of non-parametric models.
- can be used for both classification and regression.
- They are easy to understand.
- Good exploratory method to detect the influential features are in your dataset.

How Decision Trees Look like



How Decision Trees make a decision

Is it alive?

No

Yes

Alive == No

How Decision Trees make a decision

Is it alive?

No

Yes

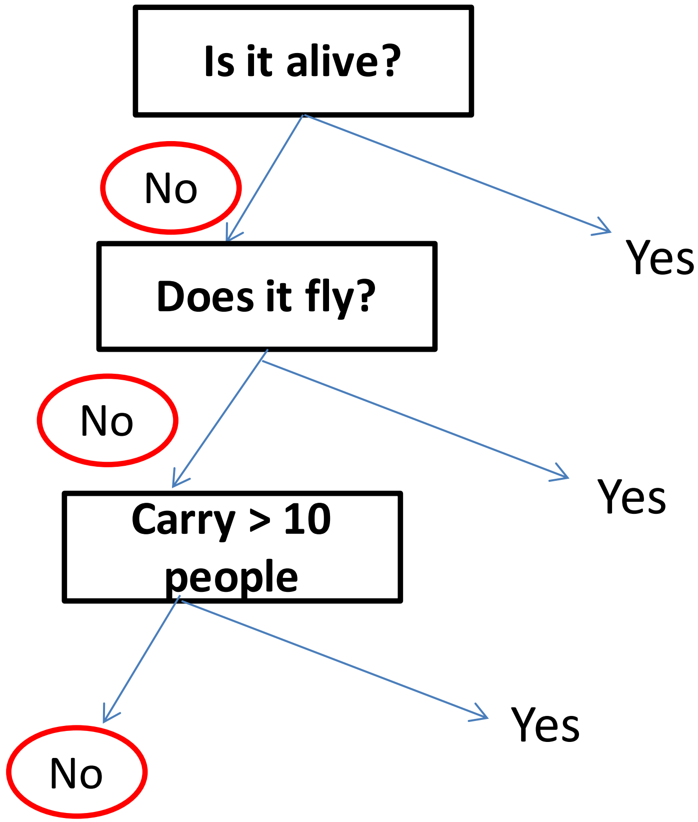
Does it fly?

No

Yes

Alive == No and fly == No

How Decision Trees make a decision



Alive == No and fly == No and carry > 10 == No

How to Determine the Best Split?

- **Gini Index**
- **Entropy**
- **Misclassification error**

Example: Gini Index

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

$$GINI_{split} = \sum_{i=1}^{k} \frac{n_i}{n} GINI(i)$$

where, n_i = number of records at child i,
n = number of records at node p.

1. Calculate

P(Hiking-labels) ?

P (Yes) = $\frac{3}{10}$, P (No) = $\frac{7}{10}$

2. Calculate P(~) All features

Table 1:

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind (F4)	Hiking (Labels)
Cloudy	Cool	Normal	Weak	No
Sunny	Hot	High	Weak	Yes
Rainy	Mild	Normal	Strong	Yes
Cloudy	Mild	High	Strong	No
Sunny	Mild	High	Strong	No
Rainy	Cool	Normal	Strong	No
Cloudy	Mild	High	Weak	Yes
Sunny	Hot	High	Strong	No
Rainy	Cool	Normal	Weak	No
Sunny	Hot	High	Strong	No

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind(F4)
P(F1 = Cloudy) = $\frac{3}{10}$	P(F2 = Cool) = $\frac{3}{10}$	P(F3=Normal) = $\frac{4}{10}$	P(F4 = Weak) = $\frac{4}{10}$
P(F1 = Sunny) = $\frac{4}{10}$	P(F2 = Hot) = $\frac{3}{10}$	P(F3 = High) = $\frac{6}{10}$	P(F4 = Strong) = $\frac{6}{10}$
P(F1 = Rainy) = $\frac{3}{10}$	P(F2 = Mild) = $\frac{4}{10}$		

Cont....Example: Gini Index

3. Get Gini Index All features to select the best root

Weather (F1)	
$P(F1 = \text{Cloudy and Hiking} = \text{Yes}) = \frac{1}{3}$	$P(F1 = \text{Cloudy and Hiking} = \text{No}) = \frac{2}{3}$
$P(F1 = \text{Sunny and Hiking} = \text{Yes}) = \frac{1}{4}$	$P(F1 = \text{Sunny and Hiking} = \text{No}) = \frac{3}{4}$
$P(F1 = \text{Rainy and Hiking} = \text{Yes}) = \frac{1}{3}$	$P(F1 = \text{Rainy and Hiking} = \text{No}) = \frac{2}{3}$

Gini Index of Cloudy = $1-((\frac{1}{3})^2+(\frac{2}{3})^2) = 0.44$

Gini Index of Sunny = $1-((\frac{1}{4})^2+(\frac{3}{4})^2) = 0.375$

Gini Index of Rainy = $1-((\frac{1}{3})^2+(\frac{2}{3})^2) = 0.44$

Weighted sum of the Gini Indices can be calculated as follow

Gini Index of Weather (F1) = $\frac{3}{10} * 0.44 + \frac{4}{10} * 0.375 + \frac{3}{10} * 0.44 = 0.414$

Temperature (F2)	
$P(F2 = \text{Cool and Hiking} = \text{Yes}) = \frac{0}{3}$	$P(F2 = \text{Cool and Hiking} = \text{No}) = \frac{3}{3}$
$P(F2 = \text{Hot and Hiking} = \text{Yes}) = \frac{1}{3}$	$P(F2 = \text{Hot and Hiking} = \text{No}) = \frac{2}{3}$
$P(F2 = \text{Mild and Hiking} = \text{Yes}) = \frac{2}{4}$	$P(F2 = \text{Mild and Hiking} = \text{No}) = \frac{2}{4}$

Gini Index of Cool = $1-((\frac{0}{3})^2+(\frac{3}{3})^2) = 0$

Gini Index of Hot = $1-((\frac{1}{3})^2+(\frac{2}{3})^2) = 0.44$

Gini Index of Mild = $1-((\frac{2}{4})^2+(\frac{2}{4})^2) = 0.5$

Weighted sum of the Gini Indices can be calculated as follows:

Gini Index of Temperature (F2) = $\frac{3}{10} * 0 + \frac{3}{10} * 0.44 + \frac{4}{10} * 0.5 = 0.332$

Table 1:

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind (F4)	Hiking (Labels)
Cloudy	Cool	Normal	Weak	No
Sunny	Hot	High	Weak	Yes
Rainy	Mild	Normal	Strong	Yes
Cloudy	Mild	High	Strong	No
Sunny	Mild	High	Strong	No
Rainy	Cool	Normal	Strong	No
Cloudy	Mild	High	Weak	Yes
Sunny	Hot	High	Strong	No
Rainy	Cool	Normal	Weak	No
Sunny	Hot	High	Strong	No

Cont....Example: Gini Index

3. Get Gini Index All features to select the best root

Humidty (F3)	
P(F3 = Normal and Hiking = Yes) = $\frac{1}{4}$	P(F3 = Normal and Hiking = No) = $\frac{3}{4}$
P(F3 = High and Hiking = Yes) = $\frac{2}{6}$	P(F3 = High and Hiking = No) = $\frac{4}{6}$

Gini Index of Normal = $1-((\frac{1}{4})^2+(\frac{3}{4})^2) = 0.375$

Gini Index of High = $1-((\frac{2}{6})^2+(\frac{4}{6})^2) = 0.44$

Weighted sum of the Gini Indices can be calculated as follows:

Gini Index of Humidty (F3) = $\frac{4}{10} * 0.375 + \frac{6}{10} * 0.44 = 0.414$

Wind(F4)	
P(F4 = Weak and Hiking = Yes) = $\frac{2}{4}$	P(F4 = Weak and Hiking = No) = $\frac{2}{4}$
P(F4 = Strong and Hiking = Yes) = $\frac{1}{6}$	P(F4 = Strong and Hiking = No) = $\frac{5}{6}$

Gini Index of Weak = $1-((\frac{2}{4})^2+(\frac{2}{4})^2) = 0.5$

Gini Index of Strong = $1-((\frac{1}{6})^2+(\frac{5}{6})^2) = 0.278$

Weighted sum of the Gini Indices can be calculated as follows:

Gini Index of Wind (F4) = $\frac{4}{10} * 0.5 + \frac{6}{10} * 0.278 = 0.367$

Table 1:

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind (F4)	Hiking (Labels)
Cloudy	Cool	Normal	Weak	No
Sunny	Hot	High	Weak	Yes
Rainy	Mild	Normal	Strong	Yes
Cloudy	Mild	High	Strong	No
Sunny	Mild	High	Strong	No
Rainy	Cool	Normal	Strong	No
Cloudy	Mild	High	Weak	Yes
Sunny	Hot	High	Strong	No
Rainy	Cool	Normal	Weak	No
Sunny	Hot	High	Strong	No

Cont....Example: Gini Index

4. Select the smallest Gini Index

Gini Index attributes or features

Weather (F1)	0.414
Temperature (F2)	0.332
Humidty (F3)	0.414
Wind (F4)	0.367

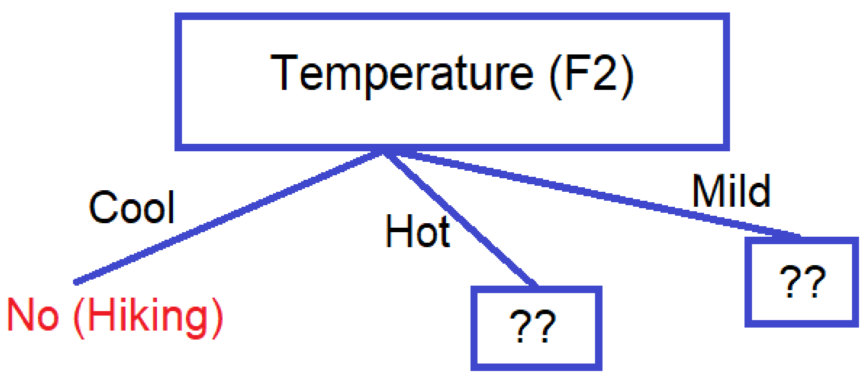


Table 1:

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind (F4)	Hiking (Labels)
Cloudy	Cool	Normal	Weak	No
Sunny	Hot	High	Weak	Yes
Rainy	Mild	Normal	Strong	Yes
Cloudy	Mild	High	Strong	No
Sunny	Mild	High	Strong	No
Rainy	Cool	Normal	Strong	No
Cloudy	Mild	High	Weak	Yes
Sunny	Hot	High	Strong	No
Rainy	Cool	Normal	Weak	No
Sunny	Hot	High	Strong	No

Cont....Example: Gini Index

Weather (F1)	
P(F1 = Sunny and Hiking = Yes) = $\frac{1}{3}$	P(F1 = Sunny and Hiking = No) = $\frac{2}{3}$

Gini Index of Sunny = $1-((\frac{1}{3})^2+(\frac{2}{3})^2) = 0.44$

Weighted sum of the Gini Indices can be calculated as follows:

Gini Index of Weather (F1) = $\frac{3}{3} * 0.44 = 0.44$

Humidty (F3)	
P(F3 = High and Hiking = Yes) = $\frac{1}{3}$	P(F3 = High and Hiking = No) = $\frac{2}{3}$

Gini Index of High = $1-((\frac{1}{3})^2+(\frac{2}{3})^2) = 0.44$

Weighted sum of the Gini Indices can be calculated as follows:

Gini Index of Humidty (F3) = $\frac{3}{3} * 0.44 = 0.44$

Wind (F4)	
P(F4 = Weak and Hiking = Yes) = $\frac{1}{1}$	
	P(F4 = Strong and Hiking = No) = $\frac{2}{2}$

Gini Index of Weak = $1-((\frac{1}{1})^2) = 0$

Gini Index of High = $1-((\frac{2}{2})^2) = 0$

Weighted sum of the Gini Indices can be calculated as follows:

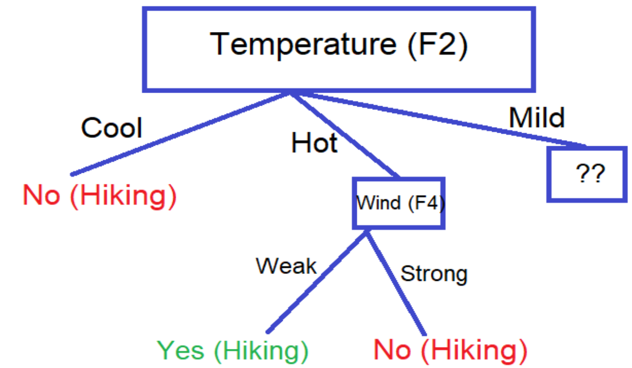
Gini Index of Wind (F4) = $\frac{1}{3} * 0 + \frac{2}{3} * 0 = 0$

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind(F4)	Hiking
Sunny	Hot	High	Weak	Yes
Sunny	Hot	High	Strong	No
Sunny	Hot	High	Strong	No

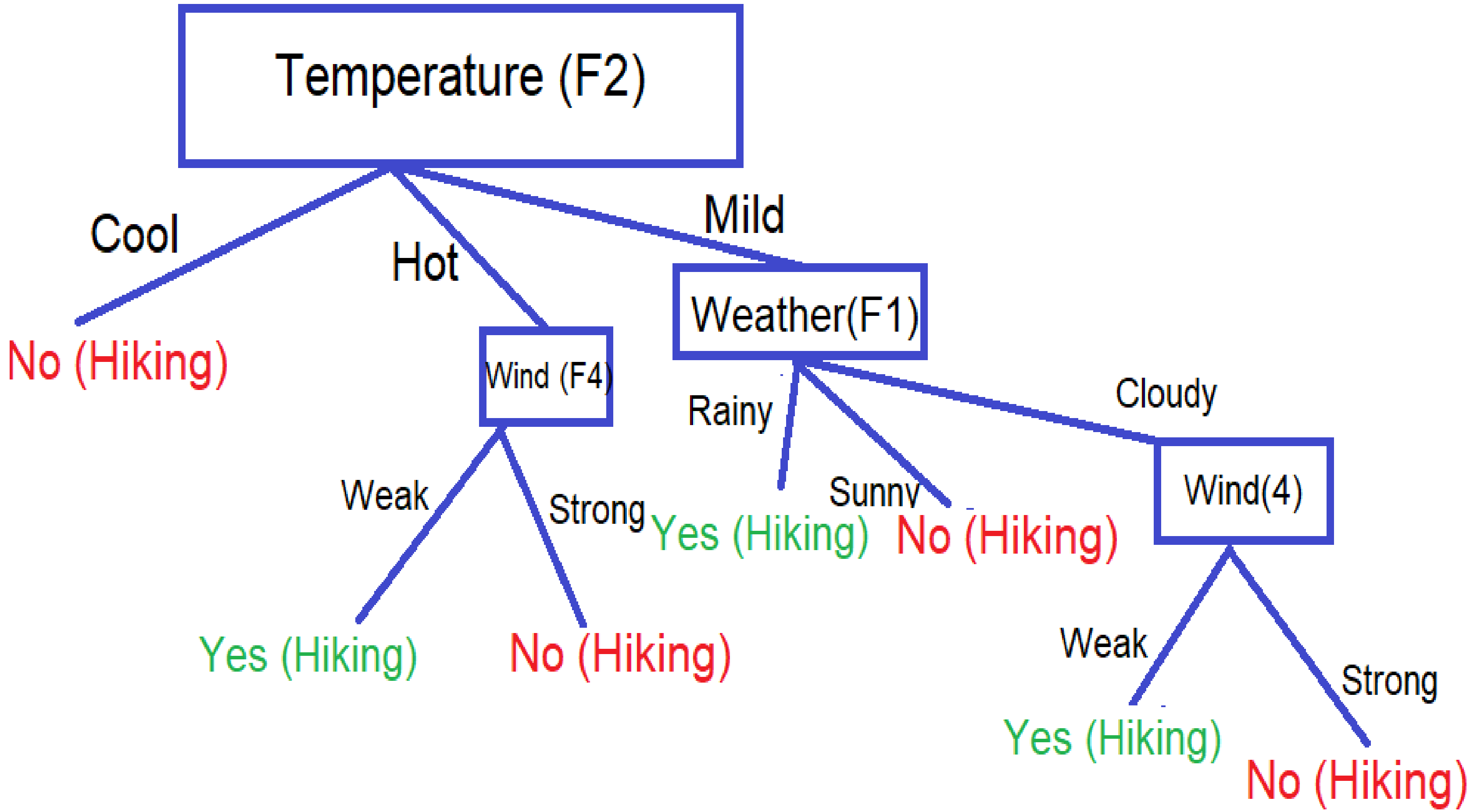
Weather (F1)	Humidty (F3)	Wind(F4)
P(F1 = Sunny) = $\frac{3}{3}$	P(F3 = High) = $\frac{3}{3}$	P(F4 = Weak) = $\frac{1}{3}$
		P(F4 = Strong) = $\frac{2}{3}$

Gini Index attributes or features

Weather (F1)	0.44
Humidty (F3)	0.44
Wind (F4)	0



Cont….Example: Gini Index



Example: *Entropy*

$$\boxed{Entropy(t) = -\sum_j p(j\,|\,t)\log_2 p(j\,|\,t)}$$

C1	**0**
C2	**6**

P(C1) = 0/6 = 0 P(C2) = 6/6 = 1

Entropy = – 0 log 0 – 1 log 1 = – 0 – 0 = 0

C1	**1**
C2	**5**

P(C1) = 1/6 P(C2) = 5/6

Entropy = $-(1/6)\log_2(1/6) - (5/6)\log_2(5/6) = 0.65$

C1	**2**
C2	**4**

P(C1) = 2/6 P(C2) = 4/6

Entropy = $-(2/6)\log_2(2/6) - (4/6)\log_2(4/6) = 0.92$

Splitting Based on Information Gain

Information Gain: Choose the split that achieves most reduction (maximizes GAIN)

$$GAIN_{split} = Entropy(p) - \left(\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right)$$

Parent Node, p is split into k partitions; n_i is number of records in partition i

Example *Information Gain* :

1. Calculate

P(Hiking-labels): P (Yes) = $\frac{3}{10}$, P (No) = $\frac{7}{10}$

Entropy (Hiking)

$= -\frac{3}{10}\ log_2(\frac{3}{10}) - \frac{7}{10}\ log_2(\frac{7}{10}) = 0.881$

2. Calculate P(~) All features

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind(F4)
P(F1 = Cloudy) = $\frac{3}{10}$	P(F2 = Cool) = $\frac{3}{10}$	P(F3=Normal) = $\frac{4}{10}$	P(F4 = Weak) = $\frac{4}{10}$
P(F1 = Sunny) = $\frac{4}{10}$	P(F2 = Hot) = $\frac{3}{10}$	P(F3 = High) = $\frac{6}{10}$	P(F4 = Strong) = $\frac{6}{10}$
P(F1 = Rainy) = $\frac{3}{10}$	P(F2 = Mild) = $\frac{4}{10}$		

Table 1:

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind (F4)	Hiking (Labels)
Cloudy	Cool	Normal	Weak	No
Sunny	Hot	High	Weak	Yes
Rainy	Mild	Normal	Strong	Yes
Cloudy	Mild	High	Strong	No
Sunny	Mild	High	Strong	No
Rainy	Cool	Normal	Strong	No
Cloudy	Mild	High	Weak	Yes
Sunny	Hot	High	Strong	No
Rainy	Cool	Normal	Weak	No
Sunny	Hot	High	Strong	No

Splitting Based on Information Gain

Cont……Example *Information Gain* :

Weather (F1)	
P(F1 = Cloudy and Hiking = Yes) = $\frac{1}{3}$	P(F1 = Cloudy and Hiking = No) = $\frac{2}{3}$
P(F1 = Sunny and Hiking = Yes) = $\frac{1}{4}$	P(F1 = Sunny and Hiking = No) = $\frac{3}{4}$
P(F1 = Rainy and Hiking = Yes) = $\frac{1}{3}$	P(F1 = Rainy and Hiking = No) = $\frac{2}{3}$

Table 1:

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind (F4)	Hiking (Labels)
Cloudy	Cool	Normal	Weak	No
Sunny	Hot	High	Weak	Yes
Rainy	Mild	Normal	Strong	Yes
Cloudy	Mild	High	Strong	No
Sunny	Mild	High	Strong	No
Rainy	Cool	Normal	Strong	No
Cloudy	Mild	High	Weak	Yes
Sunny	Hot	High	Strong	No
Rainy	Cool	Normal	Weak	No
Sunny	Hot	High	Strong	No

GAIN (Hiking, Weather (F1)) = 0.881 - $\frac{|Hiking_{Clody}|}{10}$ Entropy ($Hiking_{Clody}$) $\frac{|Hiking_{Sunny}|}{10}$ Entropy ($Hiking_{Sunny}$) $\frac{|Hiking_{Rainy}|}{10}$ Entropy ($Hiking_{Rainy}$)

$$\text{GAIN (Hiking, Weather (F1))} = 0.881 - \frac{3}{10}\left(-\frac{1}{3}\ log_2\left(\frac{1}{3}\right) - \frac{2}{3}\ log_2\left(\frac{2}{3}\right)\right) - \frac{4}{10}\left(-\frac{1}{4}\ log_2\left(\frac{1}{4}\right) - \frac{3}{4}\ log_2\left(\frac{3}{4}\right)\right) - \frac{3}{10}\left(-\frac{1}{3}\ log_2\left(\frac{1}{3}\right) - \frac{2}{3}\ log_2\left(\frac{2}{3}\right)\right)$$

= 0.881 - 0.275 - 0.234 - 0.275 = **0.097**

Temperature (F2)	
P(F2 = Cool and Hiking = Yes) = $\frac{0}{3}$	P(F2 = Cool and Hiking = No) = $\frac{3}{3}$
P(F2 = Hot and Hiking = Yes) = $\frac{1}{3}$	P(F2 = Hot and Hiking = No) = $\frac{2}{3}$
P(F2 = Mild and Hiking = Yes) = $\frac{2}{4}$	P(F2 = Mild and Hiking = No) = $\frac{2}{4}$

GAIN (Hiking, Temperature (F2))

$$= 0.881 - \frac{3}{10}\left(-\frac{0}{3}\ log_2\left(\frac{0}{3}\right) - \frac{3}{3}\ log_2\left(\frac{3}{3}\right)\right) - \frac{3}{10}\left(-\frac{1}{3}\ log_2\left(\frac{1}{3}\right) - \frac{2}{3}\ log_2\left(\frac{2}{3}\right)\right) - \frac{4}{10}\left(-\frac{2}{4}\ log_2\left(\frac{2}{4}\right) - \frac{2}{4}\ log_2\left(\frac{2}{4}\right)\right)$$

= 0.881 - 0 - 0.275 - 0.4 = **0.206**

Splitting Based on Information Gain

Cont……Example *Information Gain* :

Humidty (F3)	
P(F3 = Normal and Hiking = Yes) = $\frac{1}{4}$	P(F3 = Normal and Hiking = No) = $\frac{3}{4}$
P(F3 = High and Hiking = Yes) = $\frac{2}{6}$	P(F3 = High and Hiking = No) = $\frac{4}{6}$

$$\text{GAIN (Hiking, Humidty (F3))} = 0.881 - \frac{4}{10}\left(-\frac{1}{4}\ log_2(\frac{1}{4}) - \frac{3}{4}\ log_2(\frac{3}{4})\right) - \frac{6}{10}\left(-\frac{2}{6}\ log_2(\frac{2}{6}) - \frac{4}{6}\ log_2(\frac{4}{6})\right)$$

$$= 0.881 - 0.324 - 0.551 = \mathbf{0.006}$$

Wind(F4)	
P(F4 = Weak and Hiking = Yes) = $\frac{2}{4}$	P(F4 = Weak and Hiking = No) = $\frac{2}{4}$
P(F4 = Strong and Hiking = Yes) = $\frac{1}{6}$	P(F4 = Strong and Hiking = No) = $\frac{5}{6}$

$$\text{GAIN (Hiking, Wind (F4))} = 0.881 - \frac{4}{10}\left(-\frac{2}{4}\ log_2(\frac{2}{4}) - \frac{2}{4}\ log_2(\frac{2}{4})\right) - \frac{6}{10}\left(-\frac{1}{6}\ log_2(\frac{1}{6}) - \frac{5}{6}\ log_2(\frac{5}{6})\right)$$

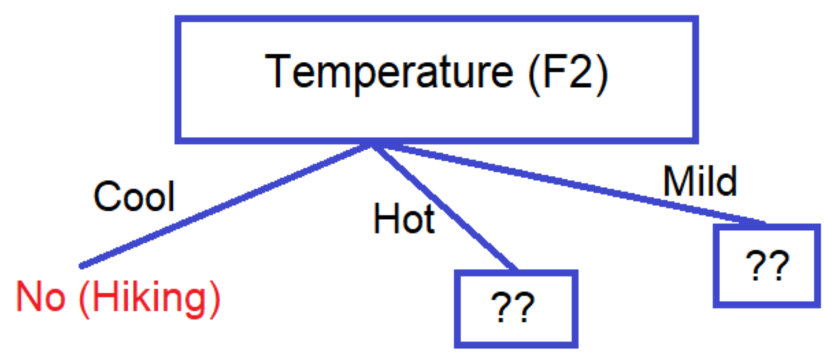
$$= 0.881 - 0.4 - 0.39 = \mathbf{0.091}$$

Information Gain attributes or features

Weather (F1)	0.097
Temperature (F2)	0.206
Humidty (F3)	0.006
Wind (F4)	0.091

Table 1:

Weather (F1)	Temperature (F2)	Humidty (F3)	Wind (F4)	Hiking (Labels)
Cloudy	Cool	Normal	Weak	No
Sunny	Hot	High	Weak	Yes
Rainy	Mild	Normal	Strong	Yes
Cloudy	Mild	High	Strong	No
Sunny	Mild	High	Strong	No
Rainy	Cool	Normal	Strong	No
Cloudy	Mild	High	Weak	Yes
Sunny	Hot	High	Strong	No
Rainy	Cool	Normal	Weak	No
Sunny	Hot	High	Strong	No



Splitting Based on Information Gain

Information Gain:

– **Disadvantage**: Tends to prefer splits that result in large number of partitions, each being small but pure.

$$Entropy(Date) = -(\frac{1}{7}log_2\frac{1}{7}+\ldots+\frac{1}{7}log_2\frac{1}{7})$$

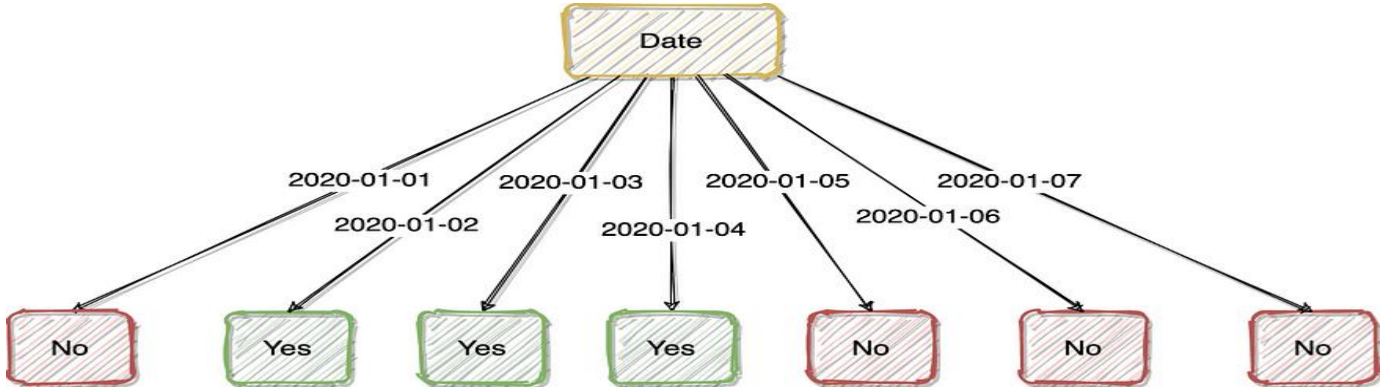
$$Entropy(Date) = -(\frac{1}{7}log_2\frac{1}{7}) \times 7$$

$$Entropy(Date) = 2.807$$

Information Gain of Weather is **0.592**
Information Gain of Temperature is **0.522**
Information Gain of Wind Level is **0.306**

Training Data

Date	Weather	Temperature	Wind Level	Go out for running?
2020-01-01	Sunny	High	Low	No
2020-01-02	Sunny	Medium	Medium	Yes
2020-01-03	Cloudy	High	Medium	Yes
2020-01-04	Cloudy	Medium	High	Yes
2020-01-05	Rainy	High	Low	No
2020-01-06	Rainy	High	Medium	No
2020-01-07	Sunny	Low	High	No



Splitting Based on Information Gain

- ***Gain Ratio:*** Designed to overcome the disadvantage of Information Gain

$$GainRATIO_{split} = \frac{GAIN_{Split}}{SplitINFO} \qquad SplitINFO = -\sum_{i=1}^{k} \frac{n_i}{n} \log \frac{n_i}{n}$$

It is simply adding a penalty on the Information Gain by dividing with the entropy of the parent node.

- ***Classification error***

$$Error(t) = 1 - \max_i P(i \mid t)$$

C1	0
C2	6

P(C1) = 0/6 = 0 P(C2) = 6/6 = 1

Error = 1 – max (0, 1) = 1 – 1 = 0

C1	1
C2	5

P(C1) = 1/6 P(C2) = 5/6

Error = 1 – max (1/6, 5/6) = 1 – 5/6 = 1/6

C1	2
C2	4

P(C1) = 2/6 P(C2) = 4/6

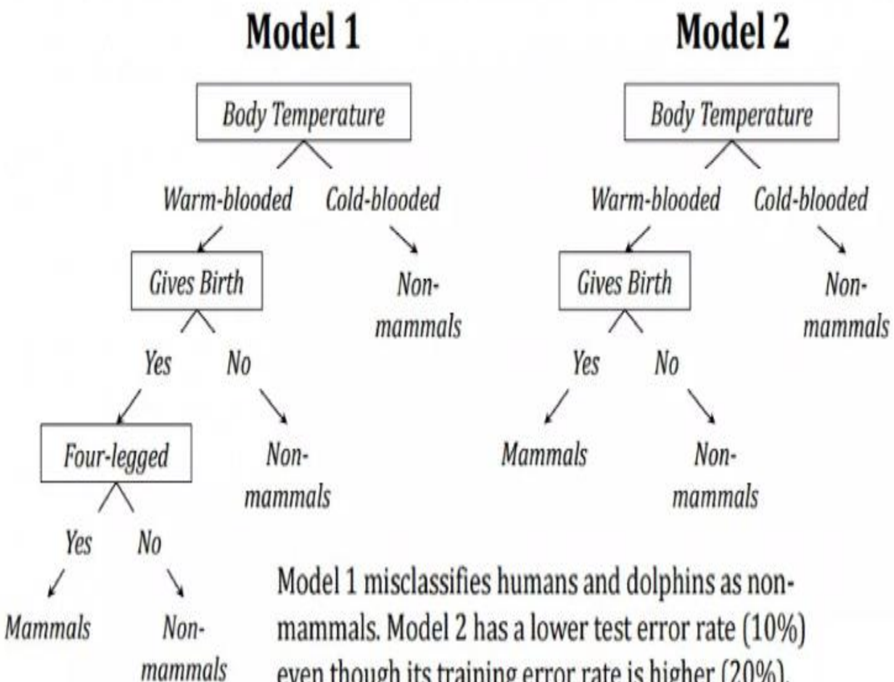
Error = 1 – max (2/6, 4/6) = 1 – 4/6 = 1/3

Issues In Decision Tree

- Overfitting Due to Noise

An example training set for classifying mammals. Asterisks denote mislabelings.

Name	Body Temperature	Gives Birth	Four-legged	Hibernates	Class Label
Porcupine	Warm-blooded	Yes	Yes	Yes	***Yes***
Cat	Warm-blooded	Yes	Yes	No	***Yes***
Bat	Warm-blooded	Yes	No	Yes	***No****
Whale	Warm-blooded	Yes	No	No	***No****
Salamander	Cold-blooded	No	Yes	Yes	***No***
Komodo dragon	Cold-blooded	No	Yes	No	***No***
Python	Cold-blooded	No	No	Yes	***No***
Salmon	Cold-blooded	No	No	No	***No***
Eagle	Warm-blooded	No	No	No	***No***
Guppy	Cold-blooded	Yes	No	No	***No***



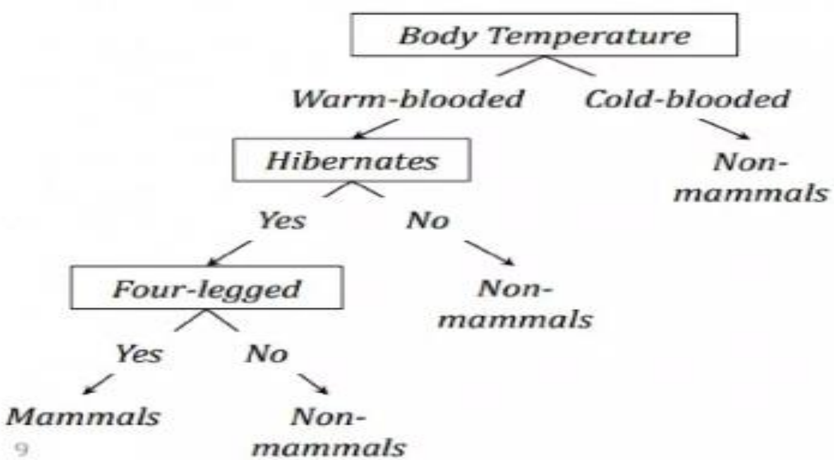
Model 1 misclassifies humans and dolphins as non-mammals. Model 2 has a lower test error rate (10%) even though its training error rate is higher (20%).

Issues In Decision Tree

- Overfitting Due to lack of Samples

An example training set for classifying mammals.

Name	Body Temperature	Gives Birth	Four-legged	Hibernates	Class Label
Salamander	Cold-blooded	No	Yes	Yes	***No***
Guppy	Cold-blooded	Yes	No	No	***No***
Eagle	Warm-blooded	No	No	No	***No***
Poorwill	Warm-blooded	No	No	Yes	***No***
Platypus	Warm-blooded	No	Yes	Yes	***Yes***



- Although the model's training error is zero, its error rate on the test set is 30%.
- Humans, elephants, and dolphins are misclassified because the decision tree classifies all warmblooded vertebrates that do not hibernate as non-mammals.

Pruning

- Pruning is a technique that controls the parts of the Decision Tree to prevent overfitting.
- There are two types of pruning: Pre-pruning and Post-pruning.
- Pre-pruntng: involves the heuristic known as ‘early stopping’ which stops the growth of the decision tree - preventing it from reaching its full depth.
- Post-pruning**:** allows the Decision Tree model to grow to its full depth by using **ccp_alphas** gives minimum leaf value of decision tree and each **ccp_aphas** will create different - different classifier and choose best out of it.

Lab

https://colab.research.google.com/drive/1YtcvzH_x397yfOdIkNfW6h5Zv8sOKqew?usp=sharing

Resources

- Stephen Marsland , "Machine Learning: An Algorithmic Perspective" second edition 2014
- https://towardsdatascience.com/decision-trees-explained-3ec41632ceb6
- https://slideplayer.com/slide/5028947/
- https://towardsdatascience.com/do-not-use-decision-tree-like-this-369769d6104d#:~:text=The%20major%20drawbacks%20of%20using,that%20has%20more%20unique%20values