

• Decision Tree Example :

toothed	hair	breathes	legs	species
True	True	True	True	Mammal
True	True	True	True	Mammal
True	false	True	false	Reptile
False	True	True	True	Mammal
True	True	True	True	Mammal
True	True	True	True	Mammal
True	false	false	false	Reptile
True	false	True	false	Reptile
True	True	True	True	Mammal
false	false	True	True	Reptile

- Find the entropy of class variable.

Here, 'species' is my class variable.

$$\therefore \text{Entropy}(\text{species}) = -p(\text{Mammal}) \cdot \log_2 p(\text{Mammal}) - p(\text{Reptile}) \cdot \log_2 p(\text{Reptile})$$

$$\begin{aligned} \text{Entropy-species} &= 6 = \text{Mammals} \\ &= 4 = \text{Reptile} \\ &= 10 = \text{total-species} . \end{aligned}$$

∴ To calculate using calculator :

$$\log_2(6/10) = \log(6/10) \div \log(2)$$

So,

$$\text{Entropy-species} = - \left[\left(\frac{6}{10} \right) \times \log \left(\frac{6}{10} \right) + \left(\frac{4}{10} \right) \times \log \left(\frac{4}{10} \right) \right]$$

~~$$= - \left[(0.6) \times \left(\frac{\log(6/10)}{\log(2)} \right) + (0.4) \times \left(\frac{\log(4/10)}{\log(2)} \right) \right]$$~~

~~$$= - \left(\frac{6}{10} \right) - (0.6) \times \left(\frac{\log(6/10)}{\log(2)} \right) - (0.4) \times \left(\frac{\log(4/10)}{\log(2)} \right)$$~~

~~$$= - (0.6)(-0.73) - (0.4)(-1.32)$$~~

$$E(s) = 0.438 + 0.528 \quad \boxed{= 0.966}$$

• Calculate Information Gain of remaining "columns/features":

- For "toothed":

toothed		Species		Total
		Mammal	Reptile	
	True	5	3	8
	False	1	1	2
				10

∴ Entropy shortcut:
 Yes/No = any value
 "0" its
 entropy is 0.
 Yes/No = same
 instances
 ↳ entropy = 1.

$$E(S, \text{toothed}) = (8/10) \times E(S, 3) + (2/10) \times E(1, 1)$$

same = "1"

$$= (8/10) \left(- (5/8) (\log_2(5/8)) - (3/8) (\log_2(3/8)) \right) + (2/10) (1)$$

$$= (0.8) \left(- (0.62) \left(\frac{\log(5/8)}{\log(2)} \right) - (0.37) \left(\frac{\log(3/8)}{\log(2)} \right) \right) + (0.2)(1)$$

$$= (0.8)(0.947) + (0.2)(1) = (0.8)(0.95) + (0.2)(1.0)$$

$$= 0.76 + 0.2$$

$$= 0.96$$

$$= 0.9635$$

$$IG(\text{toothed}) = E(S) - E(S, \text{toothed})$$

$$= 0.9666 - 0.9635$$

$$= 0.006$$

$$= 0.007$$

• For "hair":

		Mammal	Reptile	
hair	True	6	0	6
	False	0	4	4
				<hr/>
				10
				<hr/>

A brace under the '0' in the Reptile column for the 'True' row is labeled: "Because from '0' it's entropy is 0."

$$E(S, \text{hair}) = (6/10) \times E(6, 0) + (4/10) \times E(0, 4)$$

$$= (0.6) \times 0 + (0.4) \times 0.$$

$$\boxed{= 0}.$$

$$IG(\text{hair}) = 0.966 - 0.$$

$$\boxed{= 0.966}$$

• For "breather",

		Mammal	Reptile	
breathes	True	6	3	9
	False	0	1	1
				<hr/>
				10
				<hr/>

$$E(S, \text{breathes}) = (9/10) \times E(6, 3) + (1/10) \times E(0, 1).$$

$$= (0.9) \left(- (6/9) (\log_2(6/9)) - (3/9) (\log_2(3/9)) \right) + (0.1) \times 0.$$

$$= (0.9) (0.918) + 0$$

$$= 0.8264.$$

$$\boxed{= 0.82}$$

$$IG(\text{breathes}) = 0.966 - 0.82$$

$$\boxed{= 0.146}$$

• For "legs":

		Mammal	Reptile	
legs	True	6	1	7
	false	0	3	3
				<u>10</u>

$$E(s, \text{legs}) = (7/10) \times E(6, 1) + (3/10) \times E(0, 3)$$

$$= (0.7) \times \left(- (6/7) (\log_2(6/7)) - (1/7) (\log_2(1/7)) \right) + 0.3 \times 0$$

$$= (0.7)(0.59) + 0$$

$$= 0.414$$

$$IG(\text{legs}) = 0.966 - 0.414$$

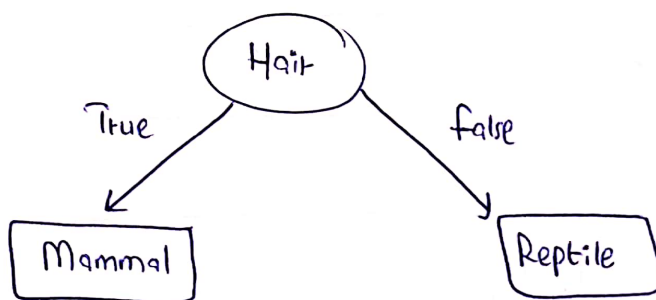
$$= 0.552$$

As,

Hair has clear side of true and false.

Hence,

It is done →



- Animal has hair ~~or not~~ : ↑

- Also we have taking decision of legs.



- We have max hair value which is ~~not~~ root node.

- Also we are creating another root node which is "legs".

↓
Continue →

• The remaining part:

legs \rightarrow True :

for "toothed":

toothed		Mammals	Reptile	Total
	True	5	0	5
	False	1	1	2
				<hr/> 7

$$E(s, \text{toothed}) = (5/7) \times E(s, 0) + (2/7) \times E(1, 1)$$

$$= (0.71) \times 0 + (0.28) \times 1$$

$$\boxed{= 0.28}$$

$$IG(\text{toothed}) = E(s) - E(s, \text{toothed})$$

$$= 0.5916 - 0.28$$

$$\boxed{= 0.3067}$$

$$E(\text{species}) = 6 \text{ mammals}$$

$$= 1 \text{ reptile}$$

$$= 7 = \text{total}$$

$$= - (6/7) (\log_2(6/7)) - (1/7) (\log_2(1/7))$$

$$\boxed{= 0.5916}$$

For "breaths":

breaths:

		Mammals	Reptile	Total
breaths	True	6	1	7
	False	0	0	0
				<hr/> 7 <hr/>

$$E(s, \overset{\text{breaths}}{\text{toothed}}) = (7/7) \times E(6,1) + (0/7) \times E(0,0)$$

$$\boxed{= .5917}$$

$$IG(\text{breaths}) = E(s) - E(s, \text{breaths}).$$

~~$$= 0.5917$$~~

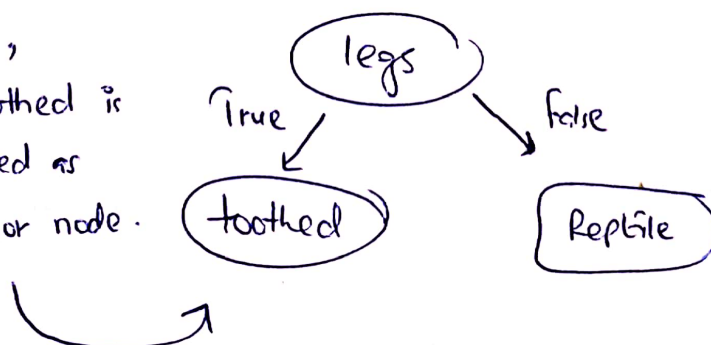
$$= 0.5917 - 0.5917$$

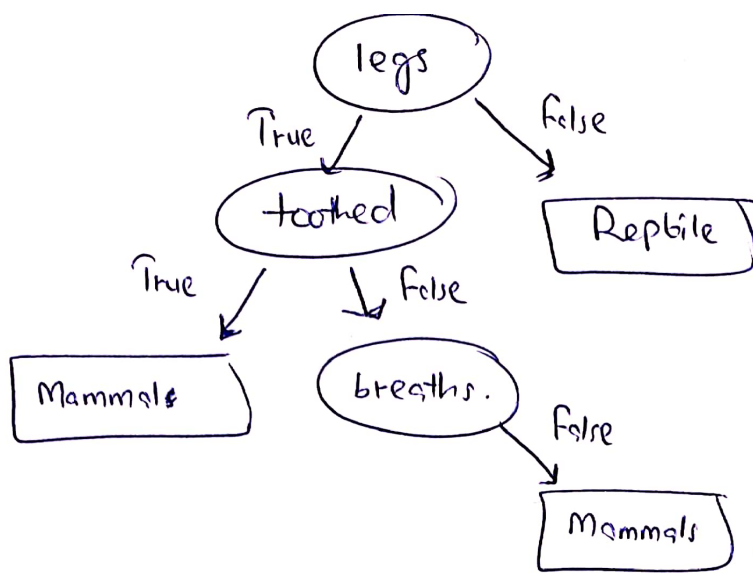
$$\boxed{= 0}$$

So, toothed has max value of IG:

Hence,

toothed is
selected as
interior node.





As,

All the true are mammals.

And false we have selected "breaths"
because it is the remaining class.

→ As we have only
one value in left
for mammals.

For "breaths"

breaths		Mammals	Reptile	total
True	1	1	0	2
False	0	0	0	0
				<hr/> 2 <hr/>

$$\begin{aligned}
 E(\text{species}) &= 1 \text{ Mammal} \\
 &= 1 \text{ Reptile} \\
 &= 2 \text{ total.}
 \end{aligned}$$

$$\begin{aligned}
 &= -\left(\frac{1}{2}\right) \left(\log_2\left(\frac{1}{2}\right)\right) - \left(\frac{1}{2}\right) \left(\log_2\left(\frac{1}{2}\right)\right) \\
 &= (-0.5)(-1) + (-0.5)(-1) \\
 &= 0.5 + 0.5
 \end{aligned}$$

$$= 1$$

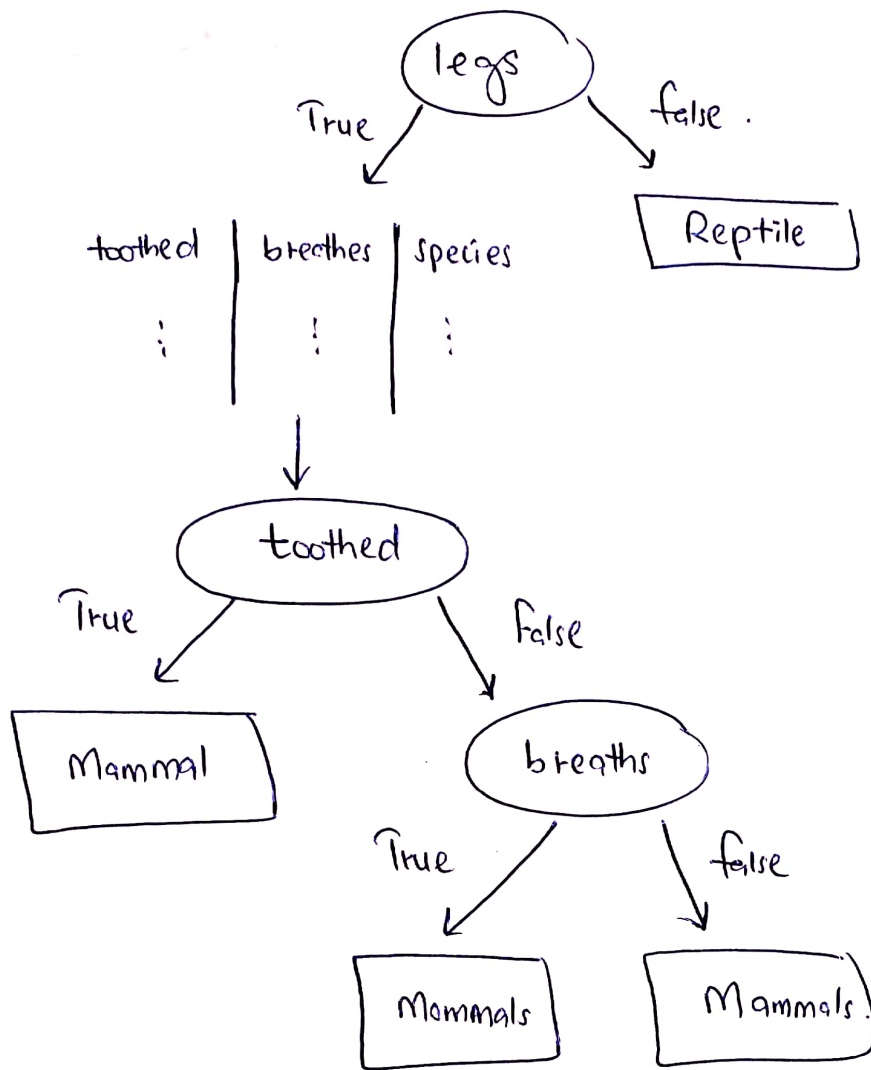
$$E(s, \text{breaths}) = \left(\frac{2}{2}\right) \times E(1, 0) + \left(\frac{0}{2}\right) \times E(0, 0)$$

$$= 1 + 0$$

$$= 1$$

$$I_f(\text{breaths}) = 1 - 1$$

$$= 0$$



- For toothed we have all are True values and false values.

Solution