Name - Pranav Vinayak Wagh
Class - TY Comp B Batch-B3
GR No. - 21810340 Roll No. - [redacted] 322061

DS SCE

(*) Table

Species	Green	Legs	Height	Smelly
M	N	3	S	Y
M	Y	2	T	N
M	Y	3	T	N
M	N	2	S	Y
M	Y	3	T	N
H	N	2	T	Y
H	N	2	S	N
H	N	2	T	N
H	Y	2	S	N
H	N	2	T	Y

Answers =>

Entropy (species) = Formula: $-\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$

P=5 N=5
(M) (H)

$$= -\frac{5}{10}\log_2\left(\frac{5}{10}\right) - \frac{5}{10}\log_2\left(\frac{5}{10}\right)$$

$$= \boxed{1}$$

→ For Green Col^n :

	M	H	Entropy
Y	3	1	0.81126
N	2	4	0.91829

↳ Entropy (Y) $= -\frac{3}{4}\log\left(\frac{3}{4}\right) - \frac{1}{4}\log\left(\frac{1}{4}\right) = \frac{0.09370 + 0.15051}{\log 2}$

$$= \frac{}{\log 2}$$

$$= 0.81126$$

↳ Entropy (N) $= -\frac{2}{6}\log\left(\frac{2}{6}\right) - \frac{4}{6}\log\left(\frac{4}{6}\right) = \frac{0.15904 + 0.11959}{\log 2}$

$$= \frac{}{\log 2}$$

$$= 0.91829$$

↳ Avg. Info (Green) = Formula = $\sum \frac{P_i + N_i}{P+N} (\text{Entropy}_{(i)})$

$= \frac{4}{10}(0.81126) + \frac{6}{10}(0.91829)$

$= 0.32450 + 0.55097$

$= 0.87547$

Answer = 2

↳ Information Gain (Green) = Entropy (Species) − Avg. Info (G)

$= 1 - 0.87547$

$= \boxed{0.12452}$

→ For Legs Col^n

	M	H	Entropy
2	2	5	0.86309
3	3	0	0

↳ Entropy (2)

$= \frac{-\frac{2}{7}\log\left(\frac{2}{7}\right) - \frac{5}{7}\log\left(\frac{5}{7}\right)}{\log 2} = \frac{0.15544 + 0.10437}{\log 2}$

$= 0.86309$

↳ Avg. Gain (Legs) $= \frac{7}{10}(0.86309) + \frac{3}{10}(0)$ (0 → 0)

$= 0.60416$

Answer = 3

↳ Info Gain (Legs) $= 1 - 0.60416$

$= \boxed{0.39583}$

→ For Height Col

	M	H	Entropy
S	2	2	1
T	3	3	1

↳ Avg Info(Height) = $\frac{4}{10}(1) + \frac{6}{10}(1)$

= 1

↳ $IG_{(Height)} = 1 - 1$

= 0

→ For Smelly Col^n

	M	H	Entropy
Y	2	2	1
N	3	3	1

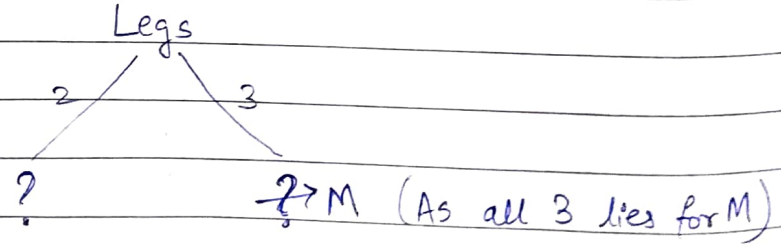
↳ Avg Info = $\frac{4}{10}(1) + \frac{6}{10}(1) = 1$

Answer = 4

↳ $IG_{(Smelly)} = 1 - 1$

= 0

→ Maximum Information Gain is of "Legs" attribute. Thus Root node is "Legs"



→ Table for Legs = 2 branch (7 Rows)

Species	Green	Legs	Height	Smelly
M	Y	2	T	N
M	N	2	S	Y
H	N	2	T	Y
H	N	2	S	N
H	N	2	T	N
H	Y	2	S	N
H	N	2	T	Y

$$\text{Entropy}_{(\text{species})} = \frac{-\frac{2}{7}\log\left(\frac{2}{7}\right) - \frac{5}{7}\log\left(\frac{5}{7}\right)}{\log 2} = \frac{0.15544 - 0.10437}{\log 2}$$

P = 2 N = 5
(M) (H)

$$= \boxed{0.86309}$$

→ For Green Col :

	M	H	Entropy
Y	1	1	1
N	1	4	0.72191

↳ $$\text{Entropy}_{(N)} = \frac{-\frac{1}{5}\log\left(\frac{1}{5}\right) - \frac{4}{5}\log\left(\frac{4}{5}\right)}{\log 2} = 0.13979 + 0.07752$$

$$= 0.72191$$

↳ $$\text{Avg Info}_{(\text{Green})} = \sum \frac{P_i + N_i}{P+N} \left(\text{Entropy}_{(i)}\right)$$

$$= \frac{2}{7}(1) + \frac{5}{7}(0.72191)$$

$$= 0.28571 + 0.51565$$

$$= 0.80136$$

↳ $IG_{(Green)} = 0.86309 - 0.80136$

$= \boxed{0.06173}$

→ For Height Col^n :

	M	H	Entropy
S	1	2	0.91829
T	1	3	0.81126

↳ $Entropy_{(S)} = \dfrac{-\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right)}{\log 2} = \dfrac{0.15904 + 0.11739}{\log 2}$

$= 0.91829$

↳ $Entropy_{(T)} = \dfrac{-\frac{1}{4}\log\left(\frac{1}{4}\right) - \frac{3}{4}\log\left(\frac{3}{4}\right)}{\log 2} = \dfrac{0.15051 + 0.09370}{\log 2}$

$= 0.81126$

$Avg\ Info = \frac{3}{7}(0.91829) + \frac{4}{7}(0.81126)$

$= 0.39355 + 0.46357$

$= 0.85712$

→ For Smelly Col^n

	M	H	Entropy
N	1	3	0.81126
Y	1	2	0.91829

} Calculations same as above.

→ $IG_{(Height)} = 0.86309 - 0.85712$

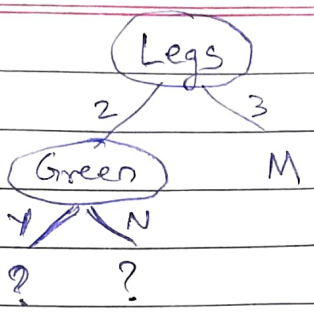
$= \boxed{0.00597}$

$Avg\ Info_{(Smelly)} = 0.85712$

$IG_{(Smelly)} = \boxed{0.00597}$

(★) Max IG is of Green Col^n = 0.06173 thus next Node is

Green



Reducing Table for Legs = 2 & Green = Y

Species	Green	Legs	Height	Smelly
M	Y	2	T	N
H	Y	2	S	N

$$\text{Entropy(species)} = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right)$$

$$= \boxed{1}$$

→ For Col^n Height

	M	H	Entropy
T	1	0	0
S	0	1	0

Avg Info = 0

IG = 1 − 0 = $\boxed{1}$

→ For Smelly Col IG = 1

(★) Thus Next Node is Height

→ Likewise we can do all the calculations for rest of tree. Finally we get a Decision tree like this :

P.T.O

⊛ Final Decision Tree