Question 1 :

① Shannon entropy of target variable (edible) is calculated as follows:

$$H(\text{edible}) = -\left(\frac{16}{24} \times \log_2\left(\frac{16}{24}\right)\right) - \left(\frac{8}{24} \times \log_2\left(\frac{8}{24}\right)\right) = 0.92 \text{ bits}$$

② Each feature white, tall, frilly's entropies & information gains are calculated as follows:

a) White :

white \ edible	True	False
0	9	5
1	7	3

$H(\text{white, edible}) = P(\text{white}=0) \times E(9,5) + P(\text{white}=1) \times E(7,3)$

entropy of white=0 ← $E(9,5)$

entropy of white=1 ← $E(7,3)$

$$P(\text{white}=0) = \frac{14}{24}, \quad E(9,5) = -\left(\frac{9}{14} \times \log_2\left(\frac{9}{14}\right)\right) - \left(\frac{5}{14} \times \log_2 \frac{5}{14}\right) = 0.94$$

$$P(\text{white}=1) = \frac{10}{24}, \quad E(7,3) = -\left(\frac{7}{10} \times \log_2\left(\frac{7}{10}\right)\right) - \left(\frac{3}{10} \times \log_2\left(\frac{3}{10}\right)\right) = 0,88$$

$$H(\text{white, edible}) = \frac{14}{24} \times 0,94 + \frac{10}{24} \times 0,88 = 0.915$$

Information gain $= H(\text{edible}) - H(\text{white, edible}) = 0,92 - 0,915 = 0,005$

b) Tall :

tall \ edible	True	false
0	6	4
1	10	14

$H(tall, edible) = P(tall=0) \cdot E(6,4) + P(tall=1) \times E(10,4)$

$E(\underbrace{6,4}_{tall=0}) = -\left(\frac{6}{10} \times \log_2 \frac{6}{10}\right) - \left(\frac{4}{10} \times \log_2\left(\frac{4}{10}\right)\right) = 0{,}97$

$E(\underbrace{10,4}_{tall=1}) = -\left(\frac{10}{14} \times \log_2 \frac{10}{14}\right) - \left(\frac{4}{14} \times \log_2\left(\frac{4}{14}\right)\right) = 0{,}861$

$P(tall=0) = 10/24$

$P(tall=1) = 14/24$

$H(tall, edible) = 0{,}906$

Info gain $= H(edible) - H(tall, edible) = 0{,}92 - 0{,}906 = 0{,}014$

c) Frilly :

Frilly \ edible	True	False
0	13	3
1	3	5

$E(\underbrace{13,3}_{frilly=0}) = -\left(\frac{13}{16} \cdot \log_2\left(\frac{13}{16}\right)\right) - \left(\frac{3}{16} \cdot \log_2\left(\frac{3}{16}\right)\right)$

$= 0.6962$

$E(3,5) = 0{,}9544$

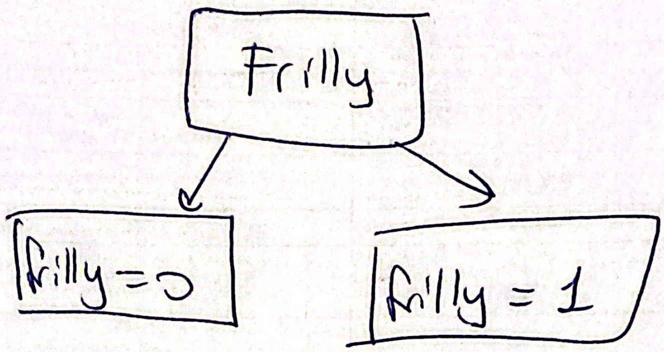
$P(frilly=0) = 16/24$

$P(frilly=1) = 8/24$

$H(frilly, edible) = 0{,}78228 \rightarrow$ Info gain $= 0{,}92 - 0{,}7822$

$= 0{,}1377$

③ Comparing information gains feature 'Frilly' should be next node with max info gain.



⇒ Frilly=0, consider attribute white's entropy.

Entropy(frilly=0) = 0,696205 as calculated before.

when frilly=0 & white=0, there are 6 True's & 3 Falses.

$$\text{entropy}(white=0) = -\frac{6}{9}\log_2\left(\frac{6}{9}\right) - \frac{3}{9}\log_2\left(\frac{3}{9}\right) = 0,92$$

entropy(white=1) = 0 since all target values are True's.

$$\text{Information gain}(frilly=0, white) = 0,696205 - \underbrace{\frac{9}{16}}_{P(white=0)} \times \underbrace{0,92}_{\text{entropy white}=0} \cancel{- \underbrace{\frac{7}{16}}_{P(white=1)} \times 0}$$

~~Ans~~

$$\text{Information gain} = 0,696205 - 0,5175 = 0,178$$

Information gain (frilly=0, white) = 0.178

$\Rightarrow$ Frilly = 0 , consider attribute tall's entropy.

Entropy (frilly=0) = 0,696205

when frilly=0 & tall=0 there are $\nearrow$ 4 True, $\searrow$ 3 False

when frilly=0 & tall=1 there are $\rightarrow$ 9 True, $\searrow$ 0 False

$$\text{entropy}(\text{tall}=0) = -\frac{4}{7}\underbrace{\log_2\left(\frac{4}{7}\right)}_{-0,81} - \frac{3}{7}\underbrace{\log_2\left(\frac{3}{7}\right)}_{-1,21} = 0,981$$

entropy (tall = 1) = 0

Information gain (frilly=0, tall) = ~~E(tall)~~

$$= 0,696205 - \underbrace{\left(\underbrace{P(\text{tall}=0)}_{\frac{7}{16}} . \underbrace{E(\text{tall}=0)}_{0,981}\right)}_{0.429} - 0$$

$= 0,2672$

(4) Therefore, Gain (white) = 0,178
Gain (~~frilly~~ tall) = 0,2672) for frilly = 0

(5) Now consider the branch frilly = 1

Scanned with CamScanner

(frilly = 1)

white	tall	edible
0	0	T
0	0	T
0	1	F
0	1	F
1	0	F
1	1	F
1	1	F
0	1	F

entropy (frilly = 1) = 0,9544

a) Consider white:

(white = 0) & frilly = 1 there are $\nearrow$ 2T $\searrow$ 3F

$$\text{entropy}(\text{white}=0) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right) = -0{,}4\times -1{,}32 - 0{,}6\times -0{,}74 = 0{,}97$$

entropy (white = 1) = 0

$$\text{Info gain}(\text{frilly}=1, \text{white}=0) = 0{,}9544 - (\underbrace{P(\text{white}=0)}_{5/8}.\underbrace{E(\text{white}=0)}_{0.97}) - 0$$

$= 0{,}35$

b) Consider tall:

tall = 0 & frilly = 1 there are $\nearrow$ 2T $\searrow$ 1F

tall = 1 & frilly = 1 there are $\nearrow$ 5F $\searrow$ 0T

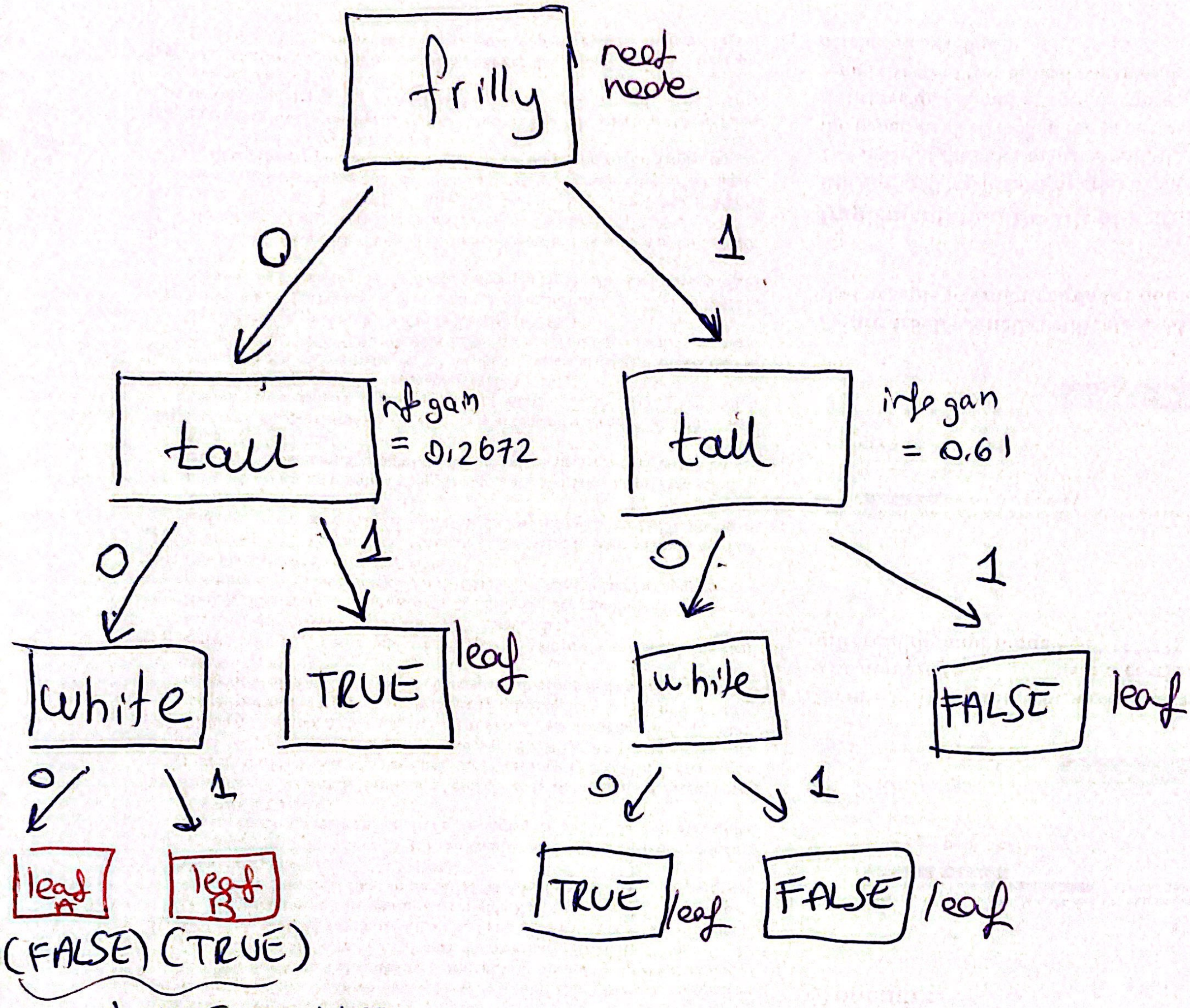
$$\text{entropy}(\text{tall}=0) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9182$$

0, 344

entropy (tall = 1) = 0 / $\text{infogain}(\text{frilly}=1, \text{tall}) = 0{,}9544 - 0{,}9182\times\frac{3}{8}$

$= 0{,}61$

Therefore the tree structure is :



↓ NOTE on this.

① When frilly=0, tall=0, white=0 there are → 1T, → 3F

② " " " , white=1 " " → 3T, → 0F

In the raw data table case ① some entries point to different classifications. In this case my opinion is classify frilly=0, tall=0, white=0 (leaf A) as false and leaf B as True. So we are ignoring the false entries in case ①.