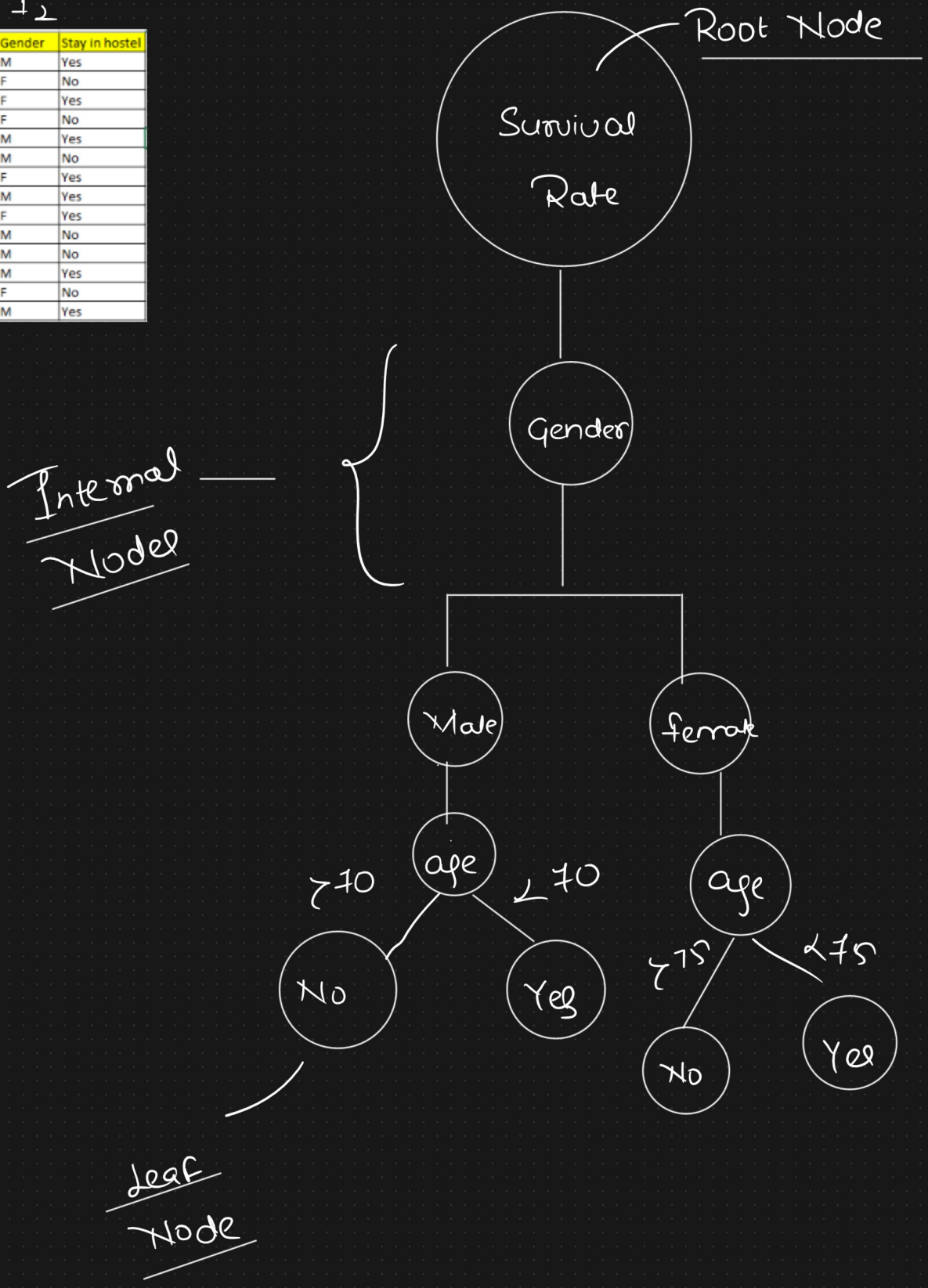
Decision Tree

f1	f2	
Class	Gender	Stay in hostel
9	M	Yes
10	F	No
8	F	Yes
8	F	No
9	M	Yes
10	M	No
11	F	Yes
11	M	Yes
8	F	Yes
9	M	No
11	M	No
11	M	Yes
10	F	No
10	M	Yes



1) Gini Index → Measure of impurity

2) Entropy

Class	Gender	Stay in hostel
9	M	Yes
10	F	No
8	F	Yes
8	F	No
9	M	Yes
10	M	No
11	F	Yes
11	M	Yes
8	F	Yes
9	M	No
11	M	No
11	M	Yes
10	F	No
10	M	Yes

class	Stay in hostel	Total
8	Y=2 N=1	3
9	Y=2 N=1	3
10	Y=1 N=3	4
11	Y=3 N=1	4
		14

n = # classes

$$1 - \sum_{i=1}^{n} (P_i)^2$$

Gini Index

Gini Index (class = 8) ⇒

$$P(Y) = 2/3 \qquad P(N) = 1/3$$

$$1 - \left(\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 \right)$$

$$= 1 - \left(\frac{4}{9} + \frac{1}{9} \right)$$

$$= 1 - 5/9 = 4/9$$

Gini Index (class = 9)

$$1 - \left(\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right)$$

$$= 4/9$$

Gini Index (class = 10)

$$1 - \left(\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2\right)$$

$$1 - \left(\frac{1}{16} + \frac{9}{16}\right)$$

$$\Rightarrow \frac{6}{16} \quad \frac{3}{8}$$

Gini Index (class = 11)

$$1 - \left(\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right)$$

$$= 3/8$$

Gini Index (class)

$$4\left(\frac{3}{14} * \frac{4}{9}\right) + \left(\frac{3}{14} * \frac{4}{9}\right) + \left(\frac{4}{14} * \frac{3}{8}\right) + \left(\frac{4}{14} * \frac{3}{8}\right)$$

$$\Rightarrow 0.404$$

Gender	stay in hostel	total
Male	Y = 5 N = 3	8
female	Y = 3 N = 3	6
		14

$$\text{Gini Index(Male)} = 1 - \left(\left(\frac{5}{8}\right)^2 + \left(\frac{3}{8}\right)^2\right)$$

$$= 1 - \left(\frac{25}{64} + \frac{9}{64}\right)$$

$$= 0.468$$

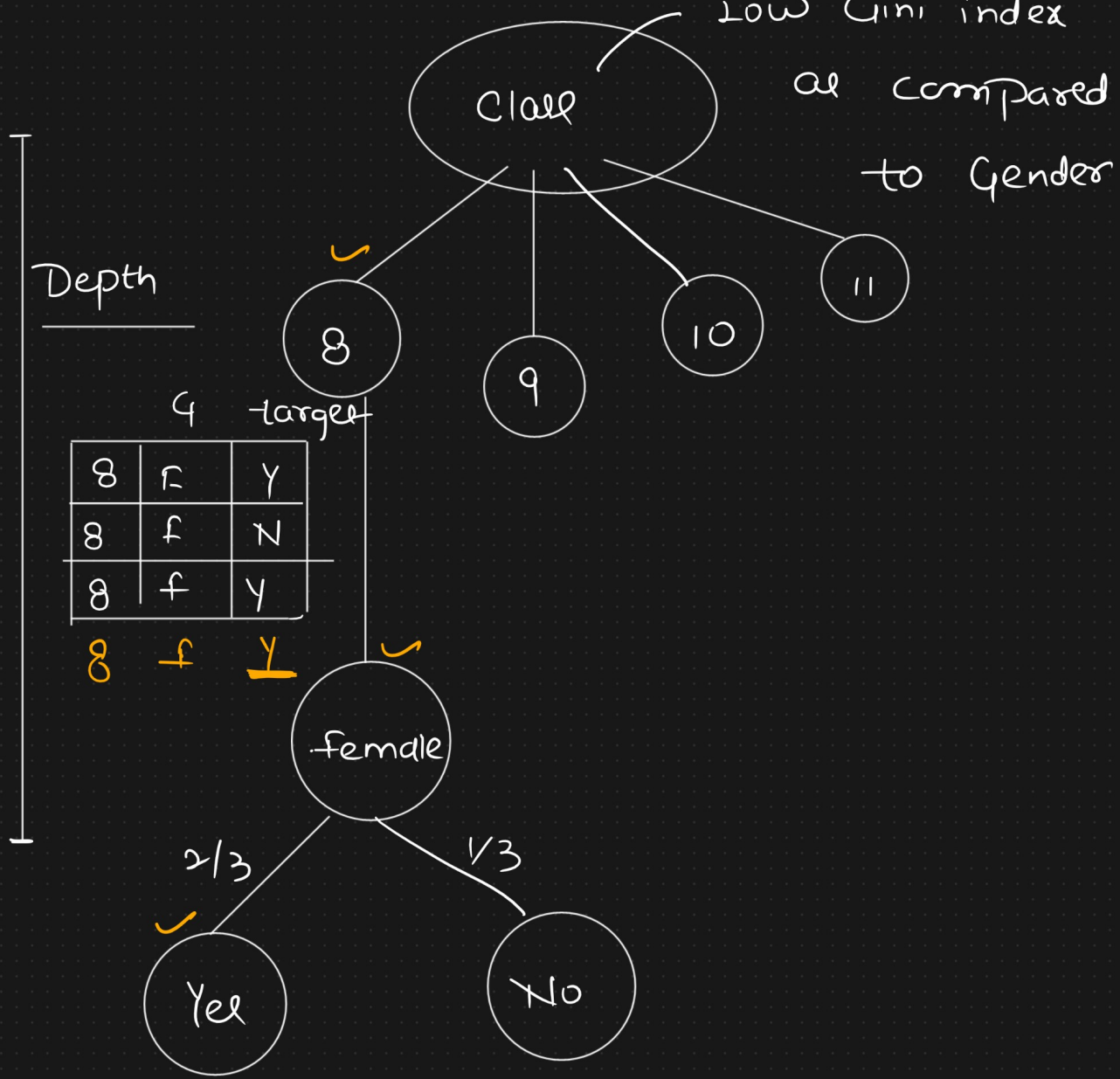
$$\text{Gini Index(female)} = 1 - \left(\left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right)$$

$$= 0.5$$

Gini Index (Gender)

↳ $\frac{8}{14} * 0.468 + \frac{6}{14} * 0.5$

$= 0.481$



Complexity

Test — Root to Leaf Node

time complexity

O(depth)

Space complexity

O(model)

Range of Gini Index (0 to 0.5)

Gender	Stay in host
f	Y = 3 N = 0

Highly Pure Sample

$$1 - \left(\left(\frac{3}{3}\right)^2 + 0 \right)$$

$$= 1 - 1 = 0$$

Gender	Stay in hostel
f	Y = 2 N = 2

$$= 1 - \left(\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2 \right)$$

$$= 1 - \left(\frac{1}{4} + \frac{1}{4}\right)$$

$$= \frac{1}{2} = \underline{\underline{0.5}}$$