Class	Stay in hostel	(n) Total values	P(Yes)	P(NO)
8	Yes = 2, NO = 1	3	2/3	1/3
9	Yes = 2, NO = 1	3	2/3	1/3
10	Yes = 1, NO = 3	4	1/4	3/4
11	Yes = 3, NO = 1	4	3/4	1/4
		M = 14		

let's Calculate Gini impurity for "class" feature

$$G_1(\text{class}=8) = 1 - [p(\text{yes})]^2 - [p(\text{no})]^2$$

$$= 1 - (2/3)^2 - (1/3)^2$$

$$= 4/9$$

$$G_1(\text{class}=9) = 1 - (2/3)^2 - (1/3)^2$$

$$= 4/9$$

$$G_1(\text{class}=10) = 1 - (1/4)^2 - (3/4)^2$$

$$= 6/16$$

$$G(\text{class} = 11) = 1 - (3/4)^2 - (1/4)^2$$

$$= 6/16$$

Weighted Sum of Gini Impurities for class features

$$G_n(\text{class}) = \frac{\text{no. of Instance(class=8)}}{\text{total Instances}} * G_i(\text{class}=8) + \frac{n(\text{class}=9)}{m} \times G(\text{class}=9)$$

(no. of Instance(class=8): let's call it 'n'; total Instances: let's call it 'm')

$$+ \frac{n(\text{class}=10)}{m} * G_i(\text{class}=10)$$

$$+ \frac{n(\text{class}=11)}{m} * G_i(\text{class}=11)$$

$$= \frac{3}{14} \times \frac{4}{9} + \frac{3}{14} \times \frac{4}{9} + \frac{4}{14} \times \frac{6}{16} + \frac{4}{14} \times \frac{6}{16}$$

$$= 0.19 + 0.214 = 0.404$$

Let's calculate Gini Impurity for "Gender" Feature

Gender	Stay on hostel	n	P(Yes)	P(NO)
M	Yes = 5, NO = 3	8	5/8	3/8
F	Yes = 3, NO = 3	6	3/6 = 1/2	1/2
		m = 14		

$$G(\text{Gender}) = \text{[illegible]} \times 1 - (5/8)^2 - (3/8)^2$$

$$= 1 - 0.39 - 0.14$$

$$G(\text{Gender} = F) = 1 - (1/2)^2 - (1/2)^2$$

$$= 0.5$$

$$G(\text{Gender}) = \frac{8}{14} \times 0.47 + \frac{6}{14} \times 0.5$$

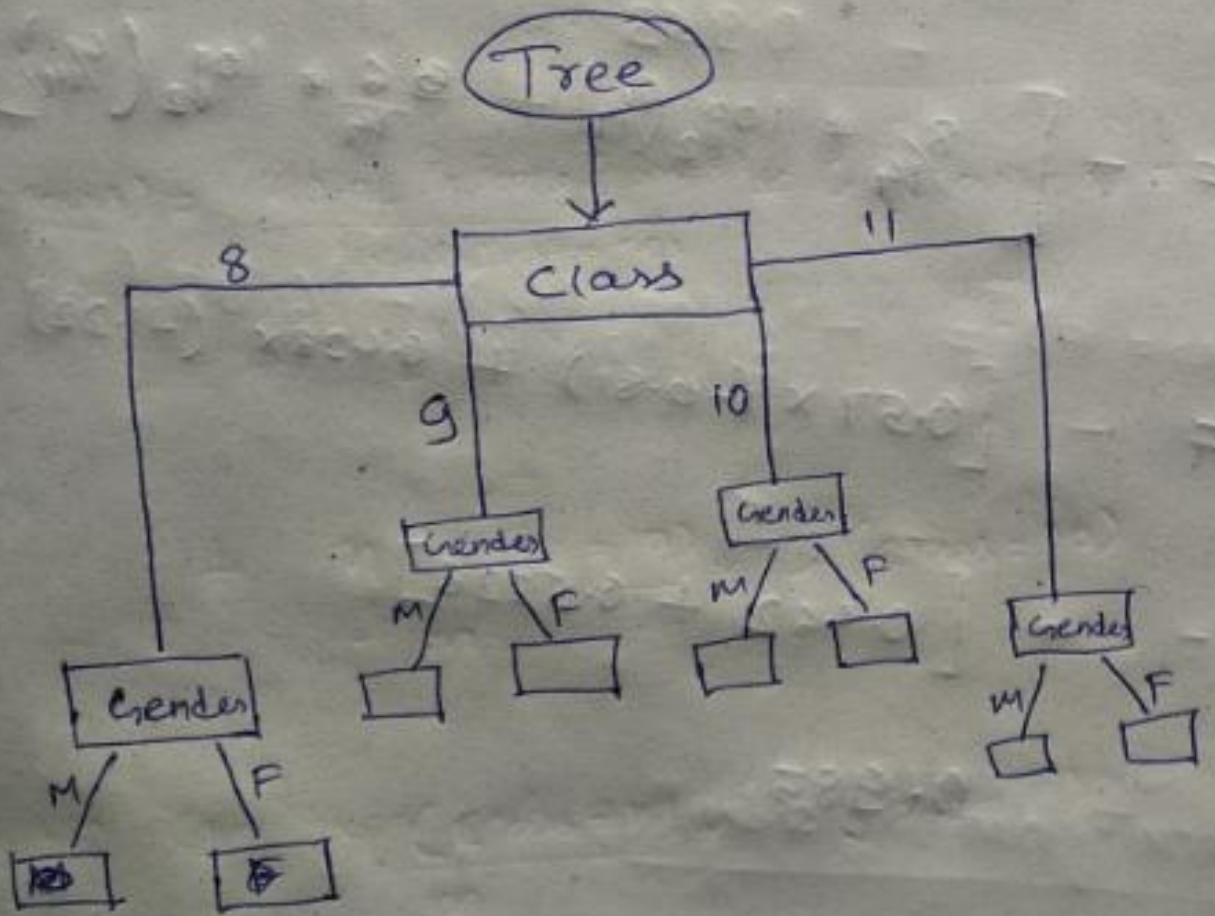
$$= 0.268 + 0.214$$

$$= 0.4822$$

We can see that,

$$G_i(\text{class}) < G_i(\text{Gender})$$

Thus, our root node is class.



Entropy of Whole Dataset

Out of 14 instances, $n(\text{Yes}) = 8$, $n(\text{NO}) = 6$

$$E(S) = -p(\text{Yes}) \log_2 p(\text{Yes}) - p(\text{NO}) \log_2 p(\text{no})$$

$$= -\left[8/14 * \log_2 (8/14) + 6/14 * \log_2 (6/14) \right]$$

$$= -\left[0.57 \times (-0.81) + 0.428 \times (-1.22) \right]$$

$$= -\left[0.462 + 0.5228 \right]$$

$$= 0.985$$

Class	Stay in hostel	(m) Total values	P(yes)	P(No)
8	Yes = 2, NO = 1	3	2/3	1/3
9	Yes = 2, NO = 1	3	2/3	1/3
10	Yes = 1, NO = 3	4	1/4	3/4
11	Yes = 3, NO = 1	4	3/4	1/4
		m = 14		

Gender	Stay on hostel	n	P(Yes)	P(NO)
M	Yes = 5, NO = 3	8	5/8	3/8
F	Yes = 3, NO = 3	6	1/2	1/2
		m = 14		

Selection of Root Node

E(class=8) = $-2/3 \log_2(2/3) - 1/3 \log_2(1/3)$

$= -[0.666(-0.586) + 0.333(-1.586)]$

$= 0.39 + 0.52 = 0.918$

E(class=9) = $-2/3 \log_2(2/3) - 1/3 \log_2(1/3)$

$= 0.918$

$$E(\text{class}=10) = -1/4 \log_2(1/4) - \frac{3}{4} \log_2(3/4)$$

$$= \text{[illegible}\ [0.5 + 0.311]$$

$$= 0.811$$

$$E(\text{class}=11) = -3/4 \log_2(3/4) - 1/4 (\log_2(1/4))$$

$$= 0.811$$

Information from "Class"

$$I(\text{class}) = \frac{3}{14} \times 0.918 + \frac{3}{14} \times 0.918 + \frac{4}{14} \times 0.811 + \frac{4}{14} \times 0.811$$

$$= 0.394 + 0.4634 = 0.8574$$

Information Gained from "Class"

$$I.G(\text{class}) = E(S) - I(\text{class})$$

$$= 0.985 - 0.857$$

$$= 0.1276$$

Genders

$$E(\text{Genders} = M) = -5/8 \log_2(5/8) - 3/8 \log_2(3/8)$$

$$= 0.423 + 0.530$$

$$= 0.953$$

$$E(\text{Genders} = F) = -1/2 \log_2(1/2) - 1/2 \log_2(1/2)$$

$$= 1$$

Information from "Gender"

$$I(\text{Gender}) = \frac{8}{14} \times 0.953 + \frac{6}{14} \times 1$$

$$= 0.544 + 0.428$$

$$= 0.972$$

Information Gain from "Gender"

$$I.G(\text{Gender}) = E(S) - I(\text{Gender})$$

$$= 0.985 - 0.972$$

$$= 0.012$$

We can clearly see,

$$I.G(\text{Gender}) < I.G(\text{class})$$

Since, "class" provides more information gain, thus our root node will be "class".