Gini Index

↳ Impurity measure

f1	f2	target
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

8 f — Y/N (Model) — Decision Tree

Lower the value of Gini Index

↳ Higher Priority (More Purity)

$$Gini(class=8) = 1 - \sum_{i=1}^{c} P_i^2$$

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

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

$$\Rightarrow 4/9$$

$$Gini(class) = \frac{3}{14} * \frac{4}{9} + \frac{3}{14} * \frac{4}{9} + \frac{4}{14} * \frac{3}{8} + \frac{4}{14} * \frac{3}{8} \Rightarrow 0.404$$

(f1)

Gini	class	Stay in Hostel	P(Y)	P(N)
4/9	← 8	Y = 2 N = 1	2/3	1/3
4/9	← 9	Y = 2 N = 1	2/3	1/3
3/8	← 10	Y = 1 N = 3	1/4	3/4
3/8	← 11	Y = 3 N = 1	3/4	1/4

$$Gini(class=10) = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2$$

$$= 1 - 1/16 - 9/16$$

$$\Rightarrow 6/16 = 3/8$$

$$Gini(Gender) = \frac{8}{14} * 0.468 + \frac{6}{14} * 0.5$$

$$\Rightarrow 0.482$$

Gender	Stay in Hostel		P(Y)	P(N)	Gini
Male	Y = 5	N = 3	5/8	3/8	0.468
female	Y = 3	N = 3	3/6	3/6	0.5

$$Gini(Male) = 1 - \left(\frac{5}{8}\right)^2 - \left(\frac{3}{8}\right)^2 = 0.468$$

$$Gini(female) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$$

Decision Tree