Ex using Gini Index

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

There are 10 samples and three classes.

Frequencies of those classes are:-

A = 3 B = 3 C = 4

$$\therefore Gi(S) = 1 - \left(\frac{3}{10}\right)^2 - \left(\frac{3}{10}\right)^2 - \left(\frac{4}{10}\right)^2$$

$Gi(S) = 0.66$; **Round 1 :- Find the root node**

1) **Attribute "Owns Home"**

For Value = yes → S_1 = 5 records ⇓ n_1 → class distribution is A = 1, B = 2 and C = 2

For Value = NO → S_2 = 5 records ⇓ n_2 → class distbn is A = 2, B = 1, C = 2

$$\therefore Gi(\text{Split on Owns-Home}) = \frac{n_1}{S} Gi(S_1) + \frac{n_2}{S} Gi(S_2)$$

$$\therefore Gi(\text{Owns-Home = yes}) \text{ i.e., } S_1 = 1 - \left(\frac{1}{5}\right)^2 - \left(\frac{2}{5}\right)^2 + \left(\frac{2}{5}\right)^2$$

$$= 0.64$$

$$\therefore Gi(S_2 \text{ OwnsHome = NO}) = 0.64$$

$$\therefore Gi(\text{Split on Owns-Home}) = \frac{5}{10} \times 0.64 + \frac{5}{10} \times 0.64$$

$$= 0.64$$

2. Attb "Married" :-

S_1 = Married = yes has A = 0, B = 1, C = 4 ; $n_1 = 5$

S_2 = Married = no; has A = 3, B = 2, C = 0; $n_2 = 5$

$$G(y) = 1 - (1/5)^2 - (4/5)^2 = 0.32$$

$$G(n) = 1 - (3/5)^2 - (2/5)^2 = 0.48$$

$$\therefore G(\text{Married}) = \frac{5}{10} \times 0.32 + \frac{5}{10} \times 0.48$$

$$= 0.40$$

3. Attb "Gender" :-

S_1 : Gender = Male has A = 0, B = 3, C = 0 ; $n_1 = 3$

S_2 : Gender = Female has A = 3, B = 0, C = 4; $n_2 = 7$

$$G(\text{Male}) = 1 - (3/3)^2 - (0)^2 = 0$$

$$G(\text{Female}) = 1 - (3/7)^2 - (4/7)^2 = 0.511$$

$$\therefore G(\text{Gender}) = \frac{3}{10} \times 0 + \frac{7}{10} \times 0.511 = 0.358$$

4. Attb "Employed" :-

S_1 : Employed = yes; has A = 3, B = 1, C = 4 ; $n_1 = 8$

S_2 : Employed = no, has A = 0, B = 2, C = 0; $n_2 = 10$

$$G(\text{yes}) = 1 - (3/8)^2 - (1/8)^2 - (4/8)^2 = 0.594$$

$$G(\text{no}) = 1 - (2/8)^2 = 0$$

$$\therefore G(\text{Employed}) = \frac{8}{10} \times 0.594 + \frac{2}{10} \times 0 = 0.475$$

5) Attb "Credit Rating"

S_1 : Credit Rating = A ; has A=2, B=1, C=2; $n_1=5$

S_2 : Credit Rating = B ; has A=1, B=2, C=2; $n_2=5$

$$G(A) = 1-\left(\frac{2}{5}\right)^2-\left(\frac{1}{5}\right)^2-\left(\frac{2}{5}\right)^2 = 0.64$$

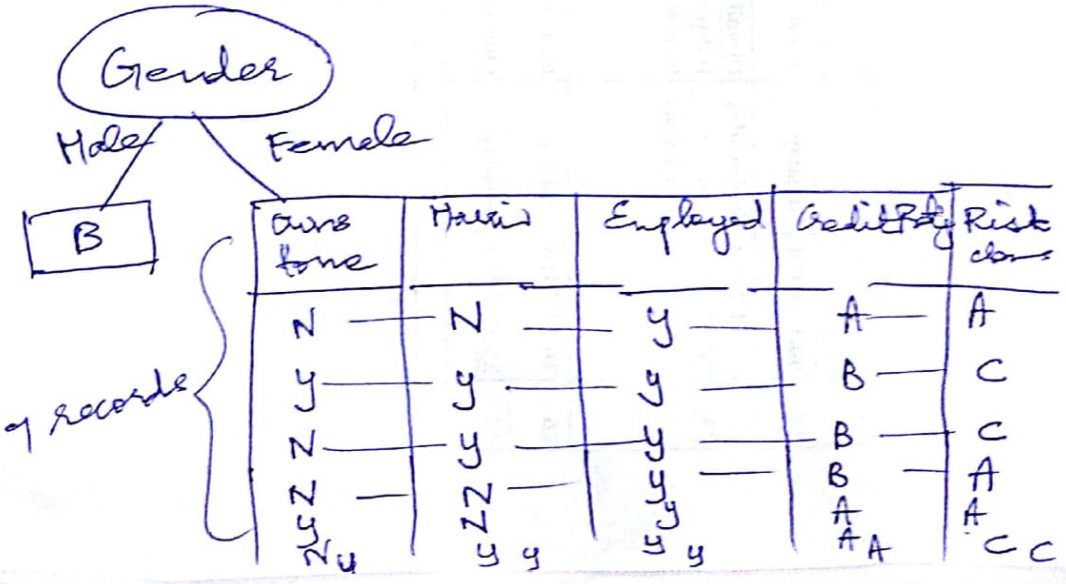
$$G(B) = 1-\left(\frac{1}{5}\right)^2-\left(\frac{2}{5}\right)^2-\left(\frac{2}{5}\right)^2 = 0.64 = G(A)$$

$$\therefore G(\text{Credit Rating}) = \frac{5}{10}\times 0.64 + \frac{5}{10}\times 0.64$$

$$= 0.64$$

Attb	Gini index before Split	Gini index after Split	Gain
Owns Home	0.66	0.64	0.02
Married	0.66	0.40	0.26
Gender	0.66	0.358	**0.302**
Employed	0.66	0.475	0.185
Credit Rating	0.66	0.64	0.02

∴ Gender is the root node



Round 2:

Owns Home	Married	Employed	Credit Rating	Risk class
N	N	y	A	A
y	y	y	B	C
N	y	y	B	C
N	N	y	B	A
y	N	y	A	A
N	y	y	A	C
y	y	y	A	C

$S = 7$; $A = 3$ $B = 0$ $C = 4$

$\therefore G(S) = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 = 0.511$

1) Attb "Owns Home":

S_1: Owns-Home = y has $A = 1$; $C = 2$; $n_1 = 3$

S_2: Owns-Home = N has $A = 2$ $C = 2$ $n_2 = 4$

$G(S_1) = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = 1 - 0.11 - 0.44$

$= 0.45$

$G(S_2) = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 1 - 0.25 - 0.25$

$= 0.5$

$\therefore G(\text{Owns-Home}) = \frac{3}{7} \times 0.45 + \frac{4}{7} \times 0.5$

$= 0.193 + 0.286 = 0.479$

2) Attb "Married":

S_1: Married = yes ; has A = 0 ; C = 4 ; $n_1 = 4$

S_2: Married = No , has A = 3 ; C = 0 , $n_2 = 3$

$G(S_1) = G(S_2) = 0$

$\therefore G(\text{Married}) = \frac{4}{7} \times 0 + \frac{3}{7} \times 0 = \underline{0}$

3) Attb "Employed":

S_1: Employed = y has A = 3 C = 4 ; $n_1 = 7$

S_2: Employed = n ; $n_2 = 0$

$\therefore G(S_1) = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 = 0.511$

$\therefore G(\text{Employed}) = \frac{7}{7} \times 0.511 = \underline{0.511}$

4) Attb "Credit Rating":

S_1: Credit Rating = A has A = 2 C = 2 $n_1 = 4$

S_2: Credit Rating = B has A = 1 C = 2 $n_2 = 3$

$G(S_1) = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$

$G(S_2) = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = 0.45$

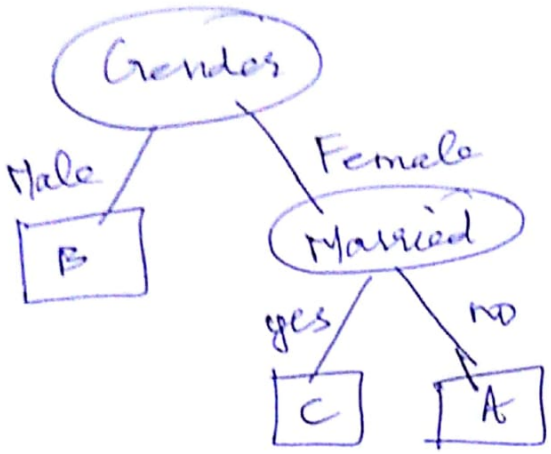
$\therefore G(\text{Credit Rating}) = \frac{4}{7} \times 0.5 + \frac{3}{7} \times 0.45$

$= 0.286 + 0.193 = \underline{0.479}$

Attb	Gini Index before split	Gini Index after split	Gain
Owns Home	0.511	0.479	0.032
Married	0.511	0	0.511
Employed	0.511	0.511	0
Credit Rating	0.511	0.479	0.032

∴ Married is the best split

∴ Tree is



← Final Tree