* Gini Index based Decision Tree ①

$$\text{Gini Index } \underline{\underline{G(s)}} = 1 - \sum_{i=1}^{n} (P_i)^2$$

- It computes the degree of probability of a specific variable that is being classified when chosen randomly & a variation of gini coefficient.

- Gini index varies between 0 to 1.

- '0' depicts that all the elements be allied to a certain class, or only one class exists there.

- The gini index of value as 1 signifies that all the elements are randomly distributed across various classes.

- A value of 0.5 denotes the elements are uniformly distributed into some classes.

* Gini Index Example ⇒

Week	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny ~~Windy~~	No	Rich	Tennis

For Gini Index, Attribute having minimum Gini Index. having the highest Information Gain (IG).

Step 1: Calculate Gini of entire dataset.

$$G(s) = 1 - \left[\left(\frac{I_{cinema}}{I_{total}}\right)^2 + \left(\frac{I_{tennis}}{I_{total}}\right)^2 + \left(\frac{I_{stay}}{I_{total}}\right)^2 + \left(\frac{I_{shopping}}{I_{total}}\right)^2\right]$$

$$G(s) = 1 - \left[\left(\frac{6}{10}\right)^2 + \left(\frac{2}{10}\right)^2 + \left(\frac{1}{10}\right)^2 + \left(\frac{1}{10}\right)^2\right]$$

$$\boxed{G(s) = 0.58}$$

Step 2: Calculate Gini Index of each attribute. ③

① $G(S_{Money}) = 1 - \left[\left(\dfrac{3}{3}\right)^2\right] = 0 \Rightarrow G(S_{Money}$

= poor $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ = poor)

$\Rightarrow$ for Money = poor, there are three
examples all with class Cinema.

For Money = Rich., 2' are having class Tennis &
1 example of Stay in & Shopping each.
3 examples of Cinema

$$G(S_{Money = Rich}) = 1 - \left[\left(\dfrac{2}{7}\right)^2 + \left(\dfrac{3}{7}\right)^2 +\right.$$

$$\left.\left(\dfrac{1}{7}\right)^2 + \left(\dfrac{1}{7}\right)^2\right.$$

$$G(S_{Money = Rich}) = 0.694$$

Weighted Average (Money)

$= G(S_{Money = Poor}) * \left(\dfrac{\text{I}_{Money = poor}}{\text{I}_{\text{total}}}\right) +$

$, G(S_{Money = Rich}) * \left(\dfrac{\text{I}_{Money = rich}}{\text{I}_{total}}\right)$

$= 0 * \left(\dfrac{3}{10}\right) + 0.694 * \left(\dfrac{2}{10}\right) = \underline{\underline{0.486}}$

(2) For Parents = Yes., all examples are with class Cinema only.

$$G(S_{Parent = Yes}) = 1 - \left(\frac{I_{Cinema_{yes}}}{I_{Parent = Yes}} \right)^2$$

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

$$= 0$$

For Parents = No, 2 examples of Tennis,
1 example of Stay In,
1 example of Shopping
1 example of Cinema.

$$G(S_{Parents = No}) = 1 - \left(\left(\frac{2}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 \right)$$

$$= 1 - \left(\overset{0.16}{0.4} + 0.0625 + 0.0625 + 0.0625 \right) 0.04$$

$$= 1 - \overset{0.04}{0.28}$$

$$= 0.72$$

Weighted Average. (Parent)

$$= 0 + \left(\frac{5}{10} \right)^2 + \underset{0.72}{0.5625} \left(\frac{5}{10} \right)$$

$$= 0.36$$

Similiary repeat for Weather.

Gini (Sunny) = $1 - \left[\left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 \right]$

$= 0.444$

Gini (Rainy) = $1 - \left[\left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 \right]$

$= 0.444$

Gini (Windy) = $1 - \left[\left(\frac{3}{4} \right)^2 + \left(\frac{1}{4} \right)^2 \right]$

$= 0.375$

∴ Weighted Average = $0.444 * \left(\frac{3}{10} \right) + 0.444 \left(\frac{3}{10} \right)$

$+ 0.375 * \left(\frac{4}{10} \right)$

$= \underline{\underline{0.416}}$

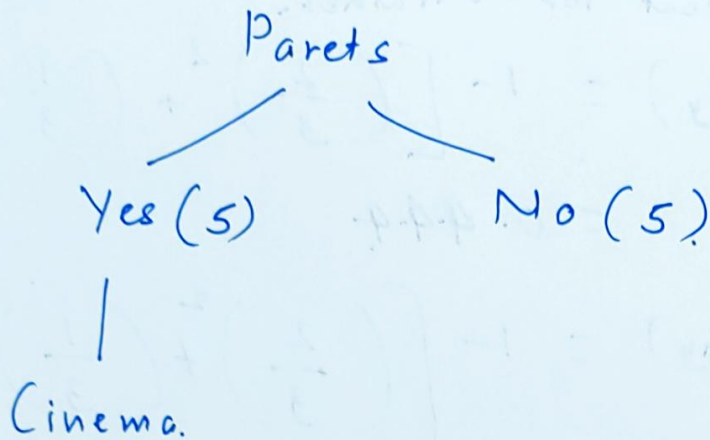For weather ⇒ Gini Index = 0.416

For Parents ⇒ Gini Index = $\underline{0.36}$

For Money ⇒ Gini Index = 0.486

∴ Here, parents having lower gini index, it will be consider as root node.

Parets

```
        Yes (5)           No (5)
          |
       Cinema.
```

For five sample where Parents is No, we will find decision attribute out of Weather & money

Five sample with Parents = No are,

Weekend	Weather	Parents	Money	Decision
W2	Sunny	No	Rich	Tennis
W5	Rainy	No	Rich	Stay In
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shoppin
W10	Sunny	No	Rich.	Tennis

For Weather, (Parents = No/ Weather)

Gini (Sunny) = 0

Gini (Rainy) = 0

Gini (Windy) = 0.5

$\therefore$ Weighted Average = $0 * \frac{2}{5} + 0 * \frac{1}{5} + 0.5 * \frac{2}{5} = 0.2$

Similary for ~~Rainy~~ Money ⇒ (Parent = No / Money)

Gini (Rich) = 0.625

Gini (Poor) = 0

Weighted Average = $\underline{\underline{0.5}}$

∴   For Parents = No / Weather ⇒ Gini Index = 0.2

For Parents = No / Weather ⇒ Gini Index = 0.5

Weather is consider as next root under for

Parent = No subtree.