

## Gini Impurity .....

Q What is Gini Impurity?

Ans Gini Impurity is same as Entropy as we discussed earlier that, Gini Impurity is also used to calculate the measure of purity / Impurity after the partition of the column. It is denoted as  $G_I$ .

Q What is different in Gini Impurity than Entropy?

Ans Formula of Entropy  $\Rightarrow \sum_{i=1}^n -p_i \log_2(p_i) \Rightarrow -P_{Yes} \log_2(P_{Yes}) - P_{No} \log_2(P_{No})$

Formula of Gini Impurity  $\Rightarrow G_I = 1 - (P_{Yes}^2 + P_{No}^2)$

Example:-

Salary	Age	Purchase
20000	21	Yes
10000	45	No
60000	27	Yes
15000	31	No
12000	18	No

Data Set - I

Salary	Age	Purchase
34000	31	No
15000	25	No
69000	57	Yes
25000	21	No
32000	28	No

Data Set - II



Date.....

$$G_I(D-I) \Rightarrow 1 - (p_{Yes}^2 + p_{No}^2)$$

$$\Rightarrow 1 - \left( \left( \frac{2}{5} \right)^2 + \left( \frac{3}{5} \right)^2 \right) \Rightarrow 1 - \left( \frac{4}{25} + \frac{9}{25} \right) \Rightarrow 1 - \frac{13}{25} \Rightarrow \boxed{0.48} \text{---(1)}$$

$$G_I(D-II) \Rightarrow 1 - (p_{Yes}^2 + p_{No}^2)$$

$$\Rightarrow 1 - \left( \left( \frac{1}{5} \right)^2 + \left( \frac{4}{5} \right)^2 \right) \Rightarrow 1 - \left( \frac{1}{25} + \frac{16}{25} \right) \Rightarrow 1 - \frac{17}{25} \Rightarrow \boxed{0.32} \text{---(2)}$$

So From (1) and (2)  $G_I(\text{DataSet-I}) > G_I(\text{DataSet-II})$

Hence Proved...