Decision Tree (Entropy)

Randomnell

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

Class	Stay in Hostel	Total
8	Y=2 N=1	3
9	Y=2 N=1	3
10	Y=1 N=3	4
11	Y=3 N=1	4

Information Gain

↳ Entropy before − Entropy after

target		Probability
Y	8	8/14
N	6	6/14

$$\text{Entropy} = -\sum_{i=1}^{n} p_i \log p_i$$

(Label)

$$= -\frac{8}{14}\log_2\frac{8}{14} - \frac{6}{14}\log_2\frac{6}{14}$$

$$= 0.98522$$

$E(8) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$ (0.918)

$E(9) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$ (0.918)

$E(10) = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.811$

$E(11) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = 0.811$

Gender	Stay in hostel	Total
Male	Y=5 N=3	8
female	Y=3 N=3	6

$E(\text{Male}) = -5/8 \log 5/8 - 3/8 \log 3/8$ → 0.954

$E(\text{female}) = -3/6 \log 3/6 - 3/6 \log 3/6$

$= 1$

$\text{Entropy (class)} = 3/14 * 0.918 + 3/14 * 0.918 + 4/14 * 0.811 + 4/14 * 0.811$

$= 0.8574$

Class

$IG = 0.98522 - 0.8574$

$= 0.1278$

$\text{Entropy(Gender)} = 8/14 * 0.954 + 6/14 * 1$

$= 0.974$

$IG(\text{Gender}) = 0.98522 - 0.974$

$= 0.01$

Higher value IG ⟶ More Priority to that node

Class ⟶ Parent Node

Y = 5 N = 5 — Highly Impure

Entropy ⇒ $-5/10 \log_2(5/10) - 5/10 \log_2(5/10)$

$= -1/2 \log_2 1/2 - 1/2 \log_2 1/2$

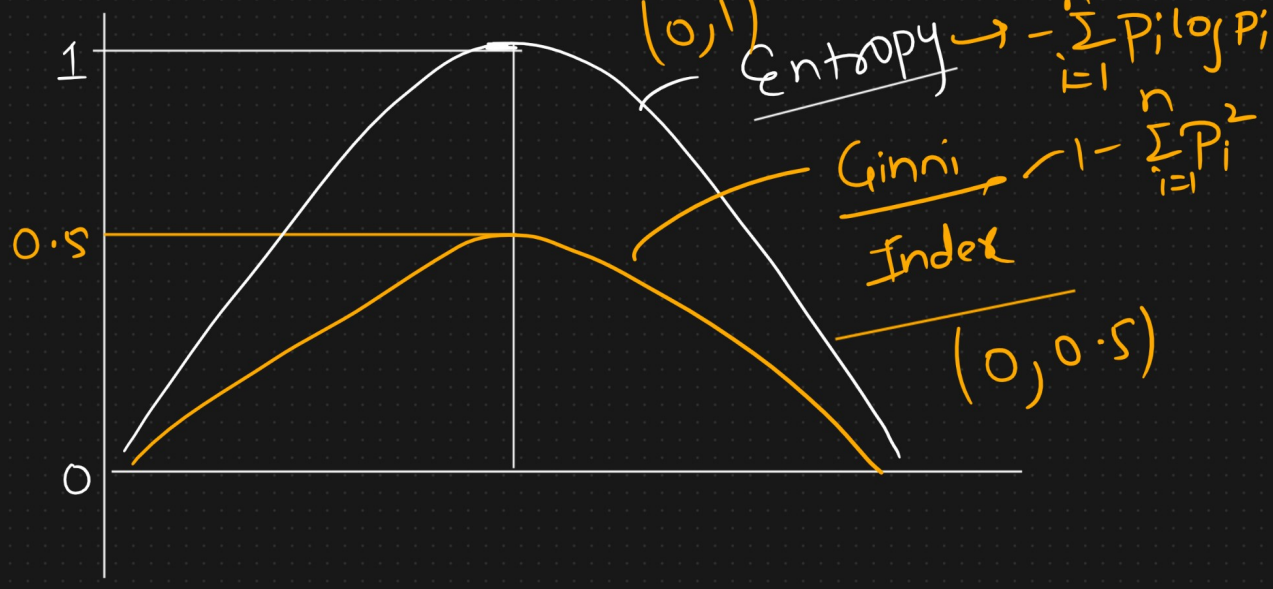
$= 1$

Highly Pure

Y = 10 N = 0

Entropy $= -10/10 \log_2 \frac{10}{10} - 0 * \log_2 0$

$= 0$


1
0.5
0
(0,1)
Entropy → $-\sum_{i=1}^{n} p_i \log p_i$
Ginni Index → $1-\sum_{i=1}^{n} p_i^2$
(0,0.5)