Date : 08-08-2025

LAB - 09 : Descision Tree

Example

Instance	a1	a2	a3	classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	yes
4	False	cool	Normal	yes
5	False	cool	Normal	yes
6	True	cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	yes
9	False	cool	Normal	yes
10	False	cool	High	yes

Solution Entropy of All Data

-> $info(D) = \frac{-6}{10} \log_2 \frac{6}{10} - \frac{4}{10} \log_2 \frac{4}{10}$

Distint val.	count
yes	6
NO	4
Total	10

$= 0.6 \times 0.7370 + 0.4 \times 1.3219$

$= 0.4422 + 0.5288$

$= 0.9710$

* Gain of a1 :

→ Gain (D, a1) = Entropy (D) - Entropy (a1)

→ Entropy of a1 :

Distinct value	yes	No	Total
True	1	4	5
False	5	0	5

→ $\text{info}_{a_1}(D) = \frac{5}{10} \times \left[\frac{-1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5} \right] \times \frac{5}{10} \times \left[\frac{-5}{5} \log_2 \frac{5}{5} \right]$

$= 0.5 \times [0.4644 + 0.2575] + 0.5 \times [1 \times 0]$

$= 0.3610$

* Gain = $\text{Info}(D) - \text{Info}_{a_1}(D)$

$= 0.9710 - 0.3610$

$= 0.61$

* Gain of a_2 :

→ Gain (D, a_2) = Entropy (D) - Entropy (a_2)

→ Entropy of a_2 :

Distinct value	yes	No	total
HOT	2	3	5
cool	4	1	5

→ $$\text{Info}_{a_2}(D) = \frac{5}{10}\left[\frac{-2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}\right]$$

$$+ \frac{5}{10}\left[\frac{-4}{5}\log_2\frac{4}{5} - \frac{1}{5}\log_2\frac{1}{5}\right]$$

$$= 0.5 \times [0.5288 + 0.4422] + 0.5 \times [0.2575 + 0.4644]$$

$$= 0.4855 + 0.3610$$

$$= 0.8465$$

→ $$\text{Gain}(a_2) = \text{Info}(D) - \text{Info}_{a_2}(D)$$

$$= 0.9710 - 0.8465$$

$$= 0.1245$$

* Gain of a_3 :

→ Gain (D, a_3) = Entropy (D) - Entropy (a_3)

→ Entropy of a_3 :

Distinct value	yes	No	Total
Normal	4	0	4
High	2	4	6

→ $\text{info}_{a_3}(D) = \frac{4}{10} \times \left[\frac{-4}{4} \log_2 \frac{4}{4} \right] + \frac{6}{10} \left[\frac{-2}{6} \log_2 \frac{2}{6} - \frac{4}{6} \log_2 \frac{4}{6} \right]$

$= [0.4 \times 0] + 0.6 [0.5283 + 0.2007]$

$= 0.4374$

→ Gain (a_3) = Info (D) - Info (a_3)

$= 0.9710 - 0.4374$

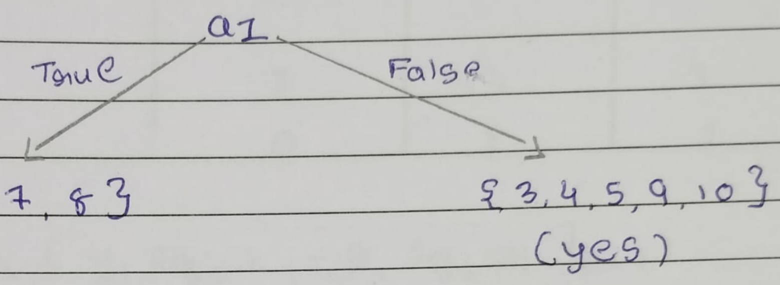
$= 0.5336$

→ Gain (a_1) = 0.61 → maximum

→ Gain (a_2) = 0.1245

→ Gain (a_3) = 0,5336

⇒



* For {1, 2, 6, 7, 8}

Instance	a_2	a_3	classification
1	Hot	High	NO
2	Hot	High	NO
6	cool	High	NO
7	Hot	High	NO
8	Hot	Normal	yes

→ Entropy of this Data

→ $info(D) = -\frac{1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5}$

$= 0,464 + 0,2575$

$= 0.7215$

Distinct value	count
yes	1
NO	4
Total	5

* Gain of a_2 :

-> Gain (D a_2) = Entropy (D) - Entropy (a_2)

-> Entropy of a_2

Distinct value	yes	No	Total
HOT	1	3	4
cool	0	1	1

-> $info_{a_2}(D) = \frac{4}{5}\left[\frac{-1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}\right]$

$+ \frac{1}{5}\left[\frac{-1}{1}\log_2\frac{1}{1}\right]$

$= 0.8 \times [0.5017 + 0.3123]$

$= 0.6512$

* Gain = $info(D) - info_{a_2}(D)$

$= 0.7215 - 0.6512$

$= 0.0703$

* Gain of a_3 :

-> Gain (D, a_3) = Entropy (D) - Entropy (a_3)

-> Entropy of a_3 :

Distinct value	yes	No	Total
High	0	4	4
Normal	1	0	1
			5

-> $$\text{info}_{a_3}(D) = \frac{4}{5}\left[\frac{-4}{4}\log_2\frac{4}{4}\right] + \frac{1}{5}\left[\frac{-1}{1}\log_2\frac{1}{1}\right]$$

$$= 0$$

* Gain $= \text{info}(D) - \text{info}_{a_3}(D)$

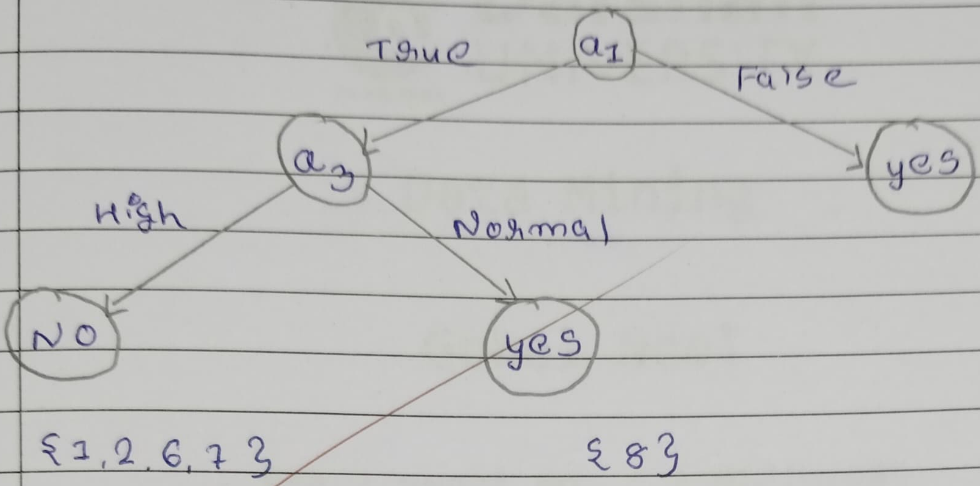
$= 0.7215 - 0$

$= 0.7215$

* Gain (a_2) = 0.0703

Gain (a_3) = 0.7215 ⟶ maximum

*


a1
True
False
a3
yes
High
Normal
NO
yes
{1,2,6,7}
{8}