Instance	a_1	a_2	a_3	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

- For Entropy of all Data

Distinct Values	Count
yes	6
No	4
Total	10

$$\text{Entropy}(D) = -\frac{6}{10}\log_2\left(\frac{6}{10}\right) - \frac{4}{10}\log_2\left(\frac{4}{10}\right)$$

$$= 0.4422 + 0.5288$$

$$= 0.9710 \text{ bits}$$

- Gain of a_1 :~

Distinct Values	Yes	No	Total
True	1	4	5
False	5	0	5
			10

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

$$= 0.5(0.7219) + 0$$

$$= 0.3609 \text{ bits}$$

$$\text{Gain}(D, a_1) = \text{Entropy}(D) - \text{Entropy}_{a_1}(D)$$

$$= 0.9710 - 0.3609$$

$$= 0.6099 \text{ bits}$$

- Gain of a_2 :-

Distinct Values	Yes	No	Total
Hot	2	3	5
Cool	4	1	5

Entropy

$$= \frac{5}{10}\left[-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] + \frac{5}{10}\left[-\frac{4}{5}\log_2\left(\frac{4}{5}\right) - \frac{1}{5}\log_2\left(\frac{1}{5}\right)\right]$$

$$= 0.8464 \text{ bits}$$

$$\text{Gain}(D, a_2) = 0.9709 - 0.8464$$

$$= 0.1245 \text{ bits}$$

- Gain of q_3 is

Distinct Values	Yes	No	Total
High	2	4	6
Normal	4	0	4

$$\text{Entropy}_{q_3}(D) = \frac{6}{10}\left[-\frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{4}{6}\log_2\left(\frac{4}{6}\right)\right] + \frac{4}{10}\left[\frac{-4}{4}\log_2\left(\frac{4}{4}\right)\right]$$

$$= 0.6\,(0.5283 + 0.3899)$$

$$= 0.5509 \text{ bits}$$

$$\text{Gain}(D, q_3) = \text{Entropy}(D) - \text{Entropy}_{q_3}(D)$$

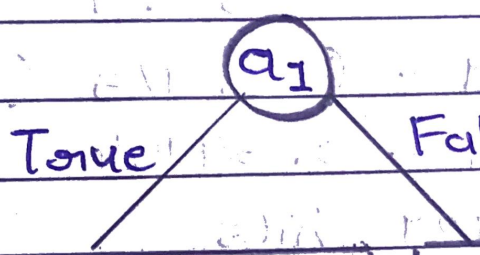
$$= 0.9710 - 0.5509$$

$$= 0.4201 \text{ bits}$$

$\Longrightarrow$ Gain $(a_1) = 0.6099 \longrightarrow$ Maximum

Gain $(a_2) = 0.1245$

Gain $(a_3) = 0.4200$



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

Instance	a_2	a_3	classification
3	Hot	High	Yes
4	Cool	Normal	Yes
5	Cool	Normal	Yes
9	Cool	Normal	Yes
10	Cool	High	Yes

- For Entropy of all Data

Distinct Values	Count
yes	1
no	4
Total	5

$$\text{Entropy}(D) = -\frac{1}{5}\log_2\left(\frac{1}{5}\right) - \frac{4}{5}\log_2\left(\frac{4}{5}\right)$$

$$= 0.4644 + 0.2575$$

$$= 0.7219 \text{ bits}$$

- Gain of q_2 :-

Distinct values	Yes	No	Total
Hot	1	3	4
Cool	0	1	1

$$\text{Gain}(D, q_2) = 0.7219 - \left[+\frac{4}{5}\left(-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}\right) + \frac{1}{5}\left(\frac{-1}{1}\log_2\frac{1}{1}\right)\right]$$

$$= 0.7219 - (0.4/5\ (0.5 + 0.3113))$$

$$= 0.7219 - 0.64904$$

$$= 0.0729 \text{ bits}$$

- Gain of q_3 :-

Distinct Values	Yes	No	Total
High	0	4	4
Normal	1	0	1

Page No. :

Date :

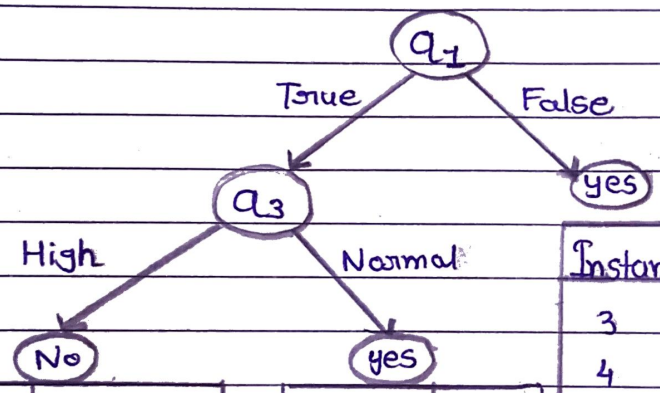
$$Gain(D, q_3) = 0.7219 - \left[\frac{4}{5}\left(-\frac{4}{4}\log_2\frac{4}{4}\right) + \frac{1}{5}\left(-\frac{1}{1}\log_2\frac{1}{1}\right)\right]$$

$$= 0.7219 - 0$$

$$= 0.7219 \text{ bits}$$

$Gain(q_2) = 0.0729$

$Gain(q_3) = 0.7219 \longrightarrow$ Maximum



Instance	q_2	classification
1	Hot	No
2	Hot	No
6	Cool	No
7	Hot	No

Instance	q_2
8	Hot
classification	
yes	

Instance	q_2	q_3	classification
3	Hot	High	yes
4	Cool	Normal	yes
5	Cool	Normal	yes
9	Cool	Normal	yes
10	Cool	High	yes