* Decision Tree :-

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 in

classification	Total
yes	6
no	4
	10

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

$$= 0.9709 \left[\because \log_y x = \frac{\log_{10} x}{\log_{10} y}\right]$$

⇒ Gain of a_1

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

Distinct values in a1	yes	no	Total
True	1	4	5
False	5	0	$\frac{5}{10}$

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

$$+$$

$$\frac{5}{10} \times \left[-\frac{5}{5} \log_2\left(\frac{5}{5}\right) - \frac{0}{5} \log_2\left(\frac{0}{5}\right) \right]$$

$$= 0.7219 \times \frac{5}{10}$$

$$= 0.3609$$

$$\therefore \text{Gain}(D, a_1) = 0.9709 - 0.3609$$
$$= 0.6099$$

⇒ Gain of a2

Distinct value in a2	yes	no	Total
Hot	2	3	5
Cool	4	1	$\frac{5}{10}$

$$Gain(D, q2) = 0.9709 - \left\{ \frac{5}{10} \left[\frac{-2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{6} \right) \right] \right.$$

Hot

$$+$$

$$\left. \frac{5}{10} \left[\frac{-4}{5} \log_2 \left(\frac{4}{5} \right) - \frac{1}{5} \log_2 \left(\frac{1}{5} \right) \right] \right\}$$

Cool

$$= 0.1245$$

$\Rightarrow$ Gain of q3

Distinct values in q3	yes	no	Total
high	2	4	6
normal	4	0	$\frac{4}{10}$

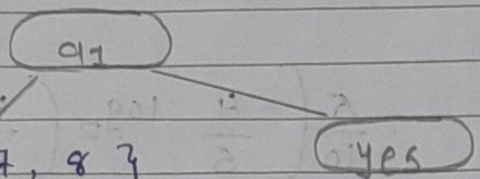
$$Gain(D, q3) = 0.9709 - \left\{ \frac{6}{10} \left[\frac{-2}{6} \log_2 \left(\frac{2}{6} \right) - \frac{4}{6} \log_2 \left(\frac{4}{6} \right) \right] \right.$$

$$+$$

$$\left. \frac{4}{10} \left[\frac{-4}{4} \log_2 \left(\frac{4}{4} \right) \right] \right\}$$

$= 0.4200$

- $Gain(D, q_1) = 0.6099$ → Maximum
  $Gain(D, q_2) = 0.1245$
  $Gain(D, q_3) = 0.4200$



- New Data

Instance	q2	q3	classification
1	Hot	High	No
2	Hot	High	No
6	Cool	High	No
7	Hot	High	No
8	Hot	Normal	Yes

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

$= 0.7219$

Dist. Value in classification	Count
yes	1
no	4
	5

⇒ Gain of q_2 :

$$\text{Gain}(D, q_2) = \text{Entropy}(D) - \text{Entropy}(q_2)$$

$$= 0.7219 - \left\{ \frac{4}{5} \left[\frac{1}{4} \log_2 \left(\frac{1}{4} \right) - \frac{3}{4} \log_2 \left(\frac{3}{4} \right) \right] \right.$$

$$+$$

$$\left. \frac{1}{5} \left[\frac{-1}{1} \log_2 \left(\frac{1}{1} \right) \right] \right\}$$

$$= 0.0729$$

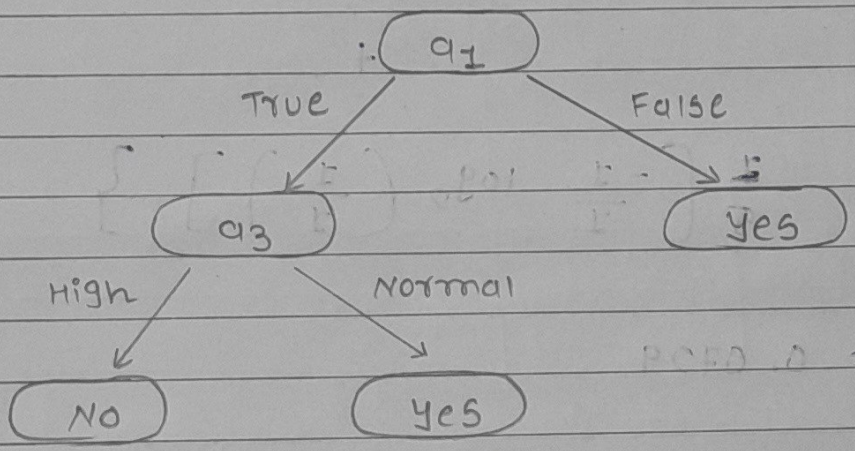
Disti. values in q_2	yes	no	total
Hot	1	3	4
Cool	1	0	$\frac{1}{5}$

⇒ Gain for q_3

Disti. value	yes	no	total
High	0	4	4
Normal	1	0	$\frac{1}{5}$

$$Gain(D, 93) = 0.7219 - \left\{ \frac{4}{5}\left[\frac{-4}{4} \log_2\left(\frac{4}{4}\right)\right] + \frac{1}{5}\left[\frac{-1}{1} \log_2\left(\frac{1}{1}\right)\right] \right\}$$

$$= 0.7219 - Max$$



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