No. 20K45A0215 Prabhu kiran. Sheet No.

Decision Tree Algorithm:

Outlook	Temparature	Humidity	Windy	hours played
Rainy	hot	high	false	25
Rainy	hot	high	True	30
Overcast	mild	high	True	52
Sunny	mild	high	false	45
Sunny	cool	high Normal	false	52
Sunny	cool	normal	True.	23
Overcast	cool	normal	True	43
Rainy	Mild	high	false	35
Rainy	cool	Normal	false	38
Rainy	Mild	Normal	True	48
Sunny	Mild	Normal	false.	46
overcast	hot	high	false	44
overcast	hot	Normal	false	44
Sunny	Mild	high	True	30

Termination criteria i.e. CV ≤ 10%.

n = 4

* standard deviation of hours played :- 9.32

calculation of standard deviation of feature columns:-

Outlook	Mean	Std. deviation	count(n)
Sunny	39.2	10.87	5
Rainy	35.2	7.78	5
Overcast	46.25	3.49	4

$SD(Overcast) = \frac{5}{14} \times 10.87 + \frac{5}{14} \times 7.78 + \frac{4}{14} \times 3.49 = 7.66$

$SDR = 1.66$

Temparature	Mean	Std. deviation	Count(n)
hot	36.25	8.95	4
cool	39	10.51	4
Mild	42.66	7.65	6

$SD(temp) = \frac{4}{14} \times 8.95 + \frac{4}{14} \times 10.5 + \frac{6}{14} \times 7.65$

$= 8.84$

$SDR = 0.48$

Humidity	Mean	std. deviation	Count(n)
High	37.57	9.36	7
Normal	42	8.73	07

$SD(Humidity) = 7/14 \times 9.36 + 7/14 \times 8.73 = 9.05$

$SDR = 0.27$

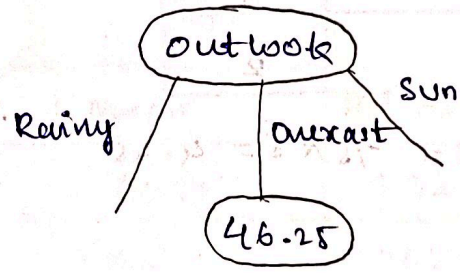
Windy	Mean	std. deviation	Count (n)
True	37.66	10.59	06
false	41.37	7.87	08

$$SD(windy) = \frac{6}{14} \times 10.59 + \frac{8}{14} \times 7.87 = 9.04$$

$$SDR = 0.28$$

→ Outlook becomes root node (high SDR)

→ for outlook overcast n≤4 and CVC=10. ∴ it will have a leaf node with output as mean of overcast five values i.e 46.25



For Outlook Rainy:-

Temp.	Humidity	Windy	hours played
Hot	high	false	25
hot	high	false True	30
Mild	high	false	35
cool	Normal	false	38
Mild	Normal	false	48

SD(hours played) = 7.78

→ calculation of std. deviation to find the next node on Outlook Rainy,

Temp.	Mean.	std. dev	count(n)
hot	27.5	2.5	02
cool	32	0	1
mild	41.5	6.5	2

SD(temp): $2/5 \times 2.5 + 1/5 \times 0 + 2/5 \times 6.5 = 3.6$

SDR = 4.18

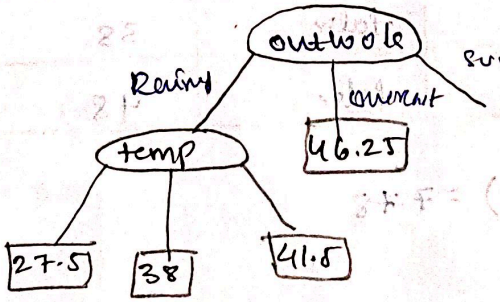
humidity	Mean	std. deviation	count
high	30	4.08	3
Normal	43	5	2

SD(humidity) = $3/5 \times 4.08 + 2/5 \times 5 = 4.45$

SDR (= 3.33

Windy	mean	std. dev.	count(n)
True	39	9	2
false	32.66	5.56	3

→ temp. has highest SDR. it becomes the next node on the Rainy branch of outlook and the leaf node is added with mean value of each attribute.



For Outlook sunny:-

Temp.	humidity	windy	hours played
Mild	high	false	45
Cool	Normal	false	52
Cool	Normal	True	23
Mild	Normal	false	46
Mild	high	True	30

SD (hours played) = 10.87.

→ Calculation of std. deviation to find the next Node on Outlook Sunny.

Temp.	Mean.	Std. deviation	Count
Mild	40.33	7.32	3
Cool	37.5	14.5	2

SD (Temp) = 3/5 × 7.32 + 2/5 × 14.5 = 10.19

SDR = 0.68

Humidity	Mean	Std. deviation	Count
high	37.5	7.5	2
Normal	40.33	12.5	3

SD(humidity) = 2/5 × 7.5 + 3/5 × 12.5 = 10.5

SDR = 0.37

windy	Mean	std dev.	Count.
True	26.5	3.5	2
false	47.67	3.09	3

→ windy have highest SDR. it becomes net node on sunny branch and all the attributes satisfy the termination criteria.

∴ leaf node are is added with mean values as o/p.