Name:- AMUDA MAHESH BABU.

PIN NO:- 20K45A0229.

Decision Tree Regression Algorithm

Outlook	Temperature	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	Normal	False	52
Sunny	Cool	Normal	False	23
Overcast	Cool	Normal	True	43
Rainy	mild	High	False	35
Rainy	Cool	Normal	False	38.
Sunny	Mild	Normal	False	46
Rainy	mild	Normal	True	48
Overcast	Hot	High	False	46
Overcast	Hot	Normal	False	44
Sunny.	mild.	High	True	30

→ Termination Criteria : CV < = 10 %.

n = 4

→ Decision tree to predict the number of hours played.

→ Standard Deviation of hours played.

S.D (Hours played) = 9.32.

* Calculation of standard deviation of feature clowns.

Outlook	Mean	Standard. Deviation	Count (n)
Sunny	39.2	10.87	5
Rainy	35.2	7.78	5
Overcast	46.25	3.49	4

$SD(Outlook) = \frac{5}{14} \times 10.87 + \frac{5}{14} \times 7.78 + \frac{4}{14} \times 3.49 = 7.6$

$SDR = 1.66.$

Temperature	Mean	Standard Deviation	Count (n)
Hot	36.25	8.95	4
Cool	39	10.51	4
Mild.	42.66	7.65	6

$SD(Temperature) = \frac{4}{14} \times 8.95 + \frac{4}{14} \times 10.51 + \frac{6}{14} \times 7.65 = 8.84$

$SDR = 0.48.$

Humidity	Mean	Standard deviation	Count (n)
High	37.57	9.36	7
Normal	42	8.73	7

$SD(Humidity) = \frac{7}{14} \times 9.36 + \frac{7}{14} \times 8.73 = 9.05$

$SDR = 0.27.$

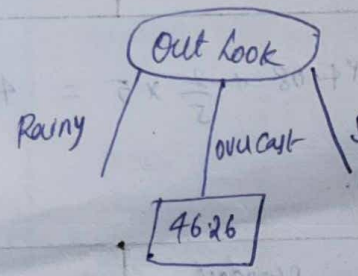
Windy	Mean	Standard deviation	Count (n)
True	37.66	10.59	6
False	41.37	7.87	8

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

$$SDR = 0.28.$$

⇒ Since out look has the higher SDR. Out Look Becomes the Root Node

⇒ For Outlook overCast, $n = 4$ and $CV = 10$ therefore it will have a leaf node with output as Mean of overCast i.e values i.e 46.26.



⇒ For Outlook Rainy :-

Temperature	Humidity	Windy	Hours played
Hot	High	False	25
Hot	High	True	30
mild	High	False	35
Cool	Normal	False	38
Mild.	Normal,	False	48.

SD(Hours played) = 7.78.

⇒ Calculation of standard deviation to find the Next Node outlook Rainy.

Temperature	Mean	Standard Deviation.	Count (n)
Hot	27.5	2.5	2
Cool	32	0	1
Mild	91.5	6.5	2

$$SD(\text{Temperature}) = \frac{2}{5} \times 2.5 + \frac{1}{5} \times 0 + \frac{2}{5} \times 6.5 = 3.6.$$

$$SDR = 4.18.$$

Humidity	Mean	Standard deviation	Count (n)
High.	30	4.08	3
Normal	43	5	2

$$SD(\text{Humidity}) = 3/5 \times 4.08 + \frac{2}{5} \times 5 = 4.45$$

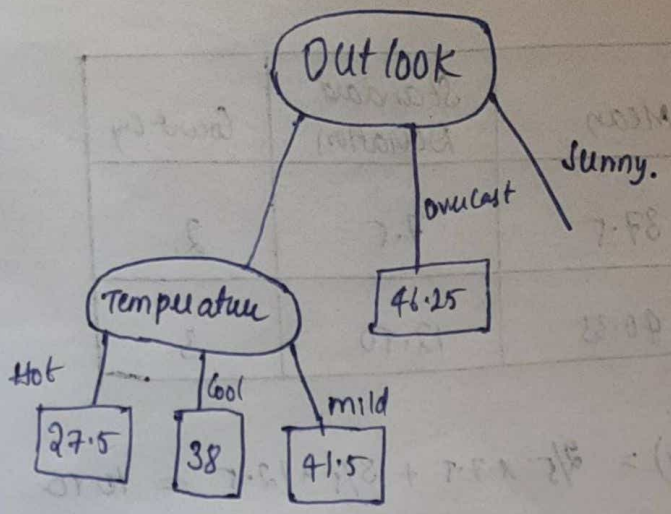
$$SDR = 3.33.$$

Windy	Mean	Standard Deviation	Count
True	39	9	2
False	32.66.	5.56	3

$$SD(\text{windy}) = 2/5 \times 9 + 3/5 \times 5.56 = 6.93.$$

$$SDR = 0.85$$

⇒ Since temperature has the heights SDR. it becomes the next node on the Rainy Branch of outlook and the leaf node are added with Mean value of Each attribute



⇒ For outlook Sunny :-

Temperature	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 Standard deviation to find the Next Node on outlook sunny.

Temperature	Mean	Standard deviation	Count (n)
Mild	40.33	7.32	3
Cool	37.5	14.5	2.

$SD(\text{temperature}) = \frac{3}{5} \times 7.32 + \frac{2}{5} \times 14.5 = 10.19.$

SDR = 0.68.

Humidity	Mean	Standard Deviation	Count
High	37.5	7.5	2
Normal	40.33	12.50	3

$SD(\text{Humidity}) = 2/5 \times 7.5 + 3/5 \times 12.5 = 10.50$

$SDR = 0.37$

Windy	Mean	Standard Deviation	Count
True	26.5	3.5	2
False	47.67	3.09	3

⇒ Since Windy has highest SDR. it Becomes Net Node on Sunny Branch and all the attributes satisfy the terminat Criteria. therefore leaf node are added with mean values as Output