Hallticket:
19K41A0464

Name:
B. Vasishth Reddy.

AI-ASSIGNMENT

Outlook	Temperature	Humidity	Windy	Hours to play
Rainy	Hot	High	False	25
Rain	Hot	High	True	30
Overcast	Hot	High	False	46
Sunny	Midd	High	False	45
Sunny	Cool	Normal	False	52
Sunny	Cool	Normal	True	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	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

⇒ Decision tree to predict no. of hours to play.

⇒ Termination Criteria : CV <= 10%,
Minimum number of samples : 4

Total Count n = 14

Average $= \bar{x} = \frac{\Sigma x}{n} = 39.8$

Standard deviation $= S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = 9.32$, (Hours played)

$CV_{(Hours played)} = \frac{S}{\bar{x}} * 100\% = 23\%$

Standard deviation of feature columns:-

Outlook:-

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

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

$SDR = 1.66$

Temperature:-

Temperature	Standard deviation	Count
Hot	8.95	4
mild	7.65	6
Cool	10.51	4

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

$SDR = 0.48$

Humidity:-

Humidity.	standard deviation	count
High	9.36	7
Normal	8.37	7.

$SD_{(Humidity)} = \frac{7}{14} \times 9.36 + \frac{7}{14} \times 8.37 = 9.05$

$SDR = 0.28$

Windy:-

Windy	Standard deviation	Count
False	7.87	8
True	10.59.	6.

$SD_{(windy)} = \frac{8}{14} \times 7.87 + \frac{6}{14} \times 10.59 = 9.02.$

$SDR = 0.29$

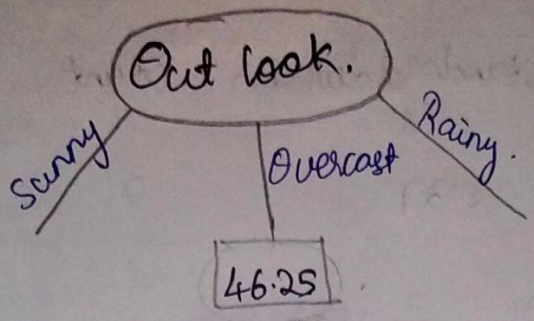
By comparing SDR of all features outlook has highest SDR, So Take Outlook as root node.

[redacted]

- Cv for Outlook:-
  - ⇒ Overcast :- 8 Y. & Count is 4
  - ⇒ Rainy :- 22 Y. & Count is 5
  - ⇒ Sunny :- 28 Y. & Count is 5

⇒ from the conditions Overcast didn't satisfied the condition. Hence it will have only a leaf node without put as mean of overcast i.e 46.25.

Tree:



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	True	48

Standard deviation(Hours played) = 7.78.

⇒ Finding the Next node :-

Temperature :-

Temperature	Standard deviation	Count(n)	Mean
Hot	2.5	2	27.5
cold	0	1	38
Mild.	6.5	2	41.5

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

$SDR = 4.18.$

Humidity:-

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

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

$SDR = 3.33$

Windy:-

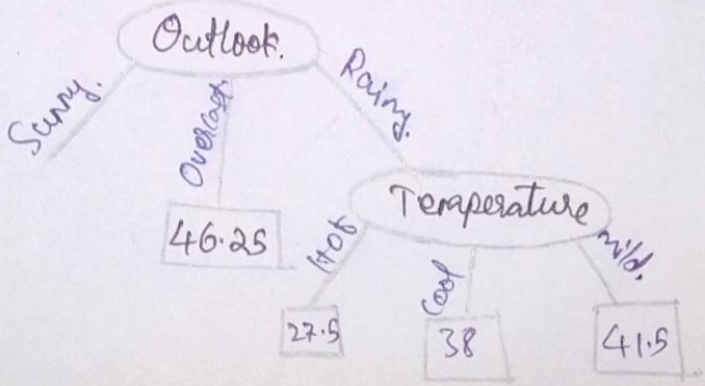
Windy	Mean	Standard Deviation	Count(n)
True	39	9	2
False	32.66	5.56	3

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

$SDR = 0.85$

⇒ Temperature has the highest SDR value. So, it becomes the next node to rainy.

⇒ The leaf nodes of Temperature are with mean value of each attribute.



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.

⇒ Finding Next node to sunny :-

Temperature :-

Temperature	Mean	Standard Deviation	(Counts)
Mild	40.33	7.32	3
Cool.	37.5	14.5	2.

SD(Temperature) :- $\frac{3}{5} \times 7.32 + \frac{2}{5} \times 14.5 = 10.19$

SDR = 0.68

Humidity:

Humidity	Mean	Standard Deviation	Count(n)
High	37.5	7.5	2
Normal.	40.33	12.50.	3.

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

$SDR = 0.37.$

Windy:

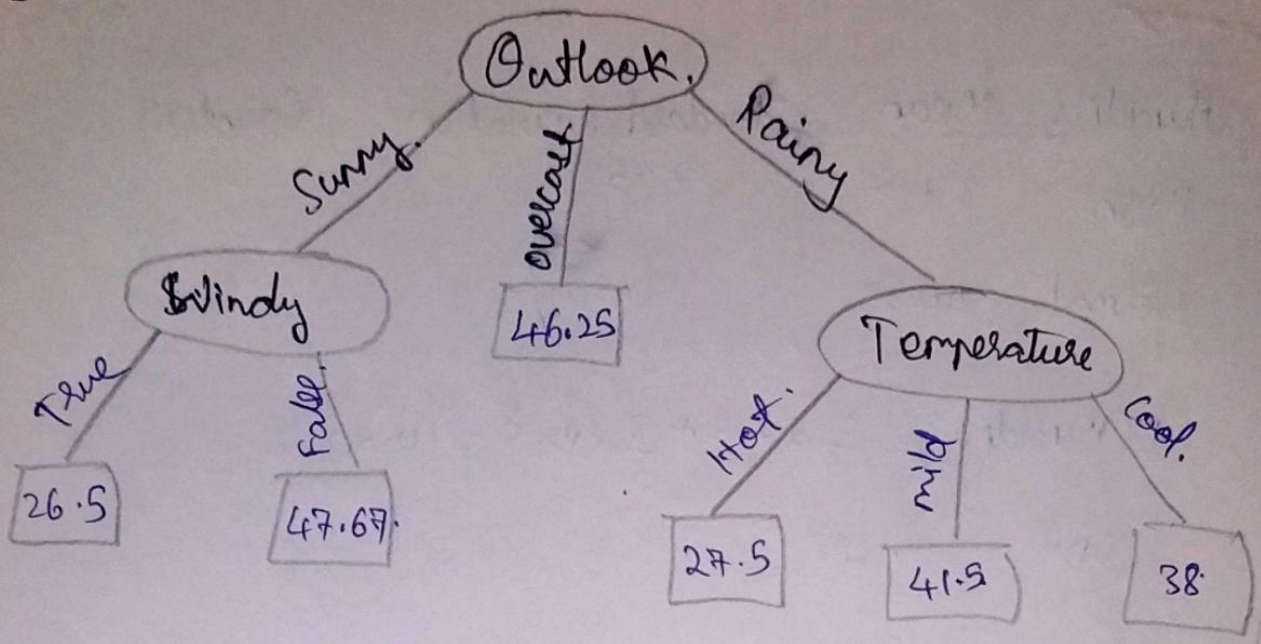
Windy	Mean	standard Deviation	Count(n).
True	26.5	3.5	2
False	47.57	3.09	3

$SD(windy) = \frac{2}{5} \times 3.5 + \frac{3}{5} \times 3.09 = 3.25$

$SDR = 7.61.$

⇒ From all the other feature windy has the highest SDR, it becomes next node on Sunny branch & the leaf nodes are added with mean values as output.

Tree:


Outlook.
Sunny.
Overcast
Rainy
Windy
46.25
Temperature
True
False.
Hot.
Mild
Cool.
26.5
47.67.
27.5
41.5
38.