ASSIGNMENT-7

596

Step 1: Read data, Identify target variable & i/p features.

Outlook	Temperature	Humidity	Windy	hours to play
Rainy	Hot	high	False	25
Rainy	Hot	high	True	30
Overcast	Hot	high	False	46
Sunny	mild	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

Step2:- Calculate SD, CN, mean.

$$\text{mean} = \frac{\Sigma x}{n}$$

$$= \frac{25+30+46+45+52+23+43+35+38+46+48+52+44+30}{14}$$

$$= \frac{557}{14} = 39.78$$

$$SD = \sqrt{\frac{\Sigma(x-\text{mean})^2}{n}}$$

$$SD = 9.67$$

$$CN = \frac{SD}{\text{mean}} \times 100 = \frac{9.67}{39.78} \times 100 = 24.30.$$

Step3:- dataset is split on different attributes the SD of each branch is calculated

$$SD(\text{attr}) = \Sigma\, w(\text{branch})\, sd(\text{branch})$$

& the result is standard deviation reduction.

$$SDR = SD - SD(\text{attr})$$

$$\therefore SD(\text{Target}) = 9.67$$

Outlook

	mean	SD	CV	n	w(v)
Rainy	35.2	8.7	24.7	5	5/14
Overcast	46.25	4.03	8.72	4	4/14
Sunny	39.2	12.2	31.0	5	5/14

$$SD(Outlook) = \frac{5}{14}(8.7) + \frac{4}{14}(4.03) + \frac{5}{14}(12.2)$$

$$= 8.59$$

$$SDR(Outlook) = SD(Target) - SD(Outlook)$$

$$= 9.67 - 8.59$$

$$= 1.08$$

Temp :-

	mean	SD	CV	n	w(v)
hot	36.25	10.34	30.6	4	4/14
Cool	39	12.14	31.1	4	4/14
mild	42.6	3.38	19.65	6	6/14

$$SD(Temp) = \frac{4}{14}(10.34) + \frac{4}{14}(12.14) + \frac{6}{14}(3.38)$$

$$= 10.01$$

$$SDR(Temp) = 9.67 - 10.01$$

$$= -0.34$$

Humidity :-

	mean	SD	CV	n	w(v)
Wgh	37.51	10.11	26.92	7	7/14
normal	42	9.4	27.4	7	7/14

$$SD(humidity) = \frac{7}{14} \times 10.11 + \frac{7}{14} \times 9.14$$

$$= 9.77$$

$$SDR(humidity) = 9.67 - 9.77$$

$$= -0.1$$

Windy :

	mean	SD	CV	n	w(CV)
True	37.6	11.6	30.8	6	6/14
False	41.3	8.41	20.3	8	8/14.

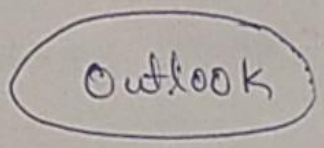
$$SD(windy) = \frac{6}{14} \times 11.6 + \frac{7}{14} \times 8.41$$

$$= 9.77$$

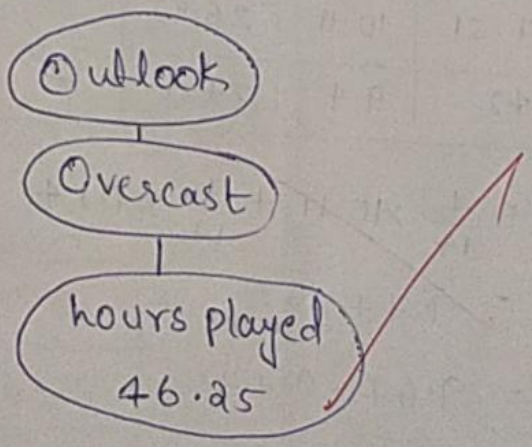
$$SDR(windy) = 9.67 - 9.77 = -0.1$$

The value that has highest SDR is considered as root node (i.e decision node)

Considering termination criteria CV is 10% or CV is ($n \leq 4$)



Overcast has CV of 8% which is less than threshold value therefore we need not to further split.



we need to split node sunny and Rainy

Outlook	Temp	humidity	windy	hours played
Sunny	mild	high	false	45
Sunny	cool	normal	false	52
sunny	cool	normal	True	23
Sunny	mild	normal	False	46
Sunny	mild	high	True	30.

mean = 39.2

SD = 12.2

CV = 31.0.

Temp:

	mean	SD	CV	n	w(v)
mild	40.3	8.96	22.23	3	3/5
cool	37.5	20.50	54.66	2	2/5

$SD(Temp) = \frac{3}{5}(8.96) + \frac{2}{5}(20.50) = 13.576$

$= 12.2 - 13.576$

$= -1.37$

humid:-

	mean	SD	CV	n	w(v)
high	37.5	10.6	28.26	2	2/5
normal	40.3	15.30	37.96	3	3/5

$SD(humid) = \frac{2}{5}(10.6) + \frac{3}{5}(15.30) +$

$= 6.4\,(10.6) + 0.6(15.30)$

$= 13.42$

$SD(humid) = 12.2 - 13.42$

$= -1.22$

windy:

	mean	SD	CV	n	w(v)
false	47.66	3.78	7.94	3	3/5
True.	26.5	4.94	18.65	2	2/5

$SD(windy) = 3/5(3.78) + \frac{2}{5}(4.94)$

$= 4.23$

$SD(windy) = 12.2 - 4.23$

$= 7.97$

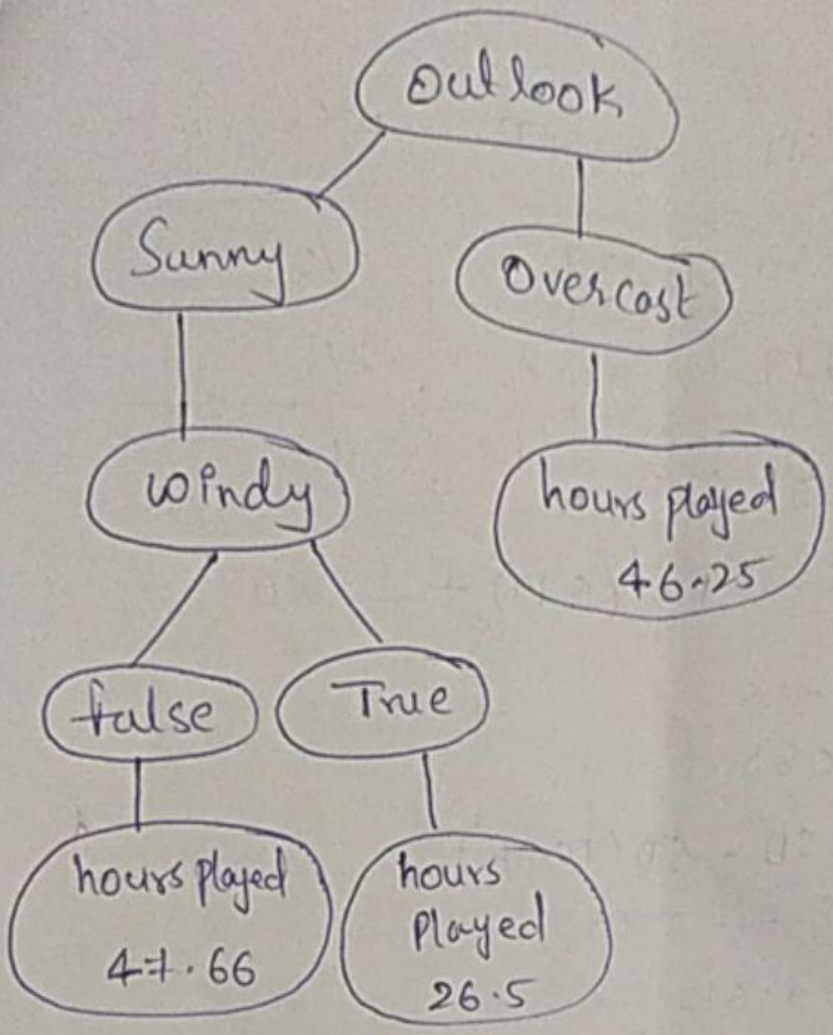
then check for highest SDR

In outlook, among Temp, humidity and windy SDR value is high for windy.

$SDR = 7.97$

Then, check for CV value.

both True & false satisfy the CV value.



Rainy :

Outlook	Temperature	humidity	windy	hours to play
Rainy	hot	high	false	25
Rainy	hot	high	True	30
Rainy	mild	high	False	35
Rainy	cool	normal	false	38
Rainy	mild	normal	True	48

∴ mean = 35.2

SD = 8.7

CV = 24.7

Temperature.

Temperature	mean	SD	CV	n	w(v)
hot	27.5	3.53	12.83	2	2/5
mild	41.5	9.19	22.144	2	2/5
cool	38	0	0	1	1/5

$$SD(Temp) = \frac{2}{5}(3.53) + \frac{2}{5}(9.19) + \frac{1}{5} \times 0$$

$$= 8.4 - 5$$

$$= 5.088$$

$$SDR(Temp) = SD - SD(Temp)$$

$$= 8.7 - 5.088$$

$$= 3.612$$

humidity :-

humidity	mean	SD	CV	n	w(v)
high	30	5	16.66	3	3/5
normal	43	7.07	16.44	2	2/5

$$SD(humidity) = \frac{3}{5}(5) + \frac{2}{5}(7.07)$$

$$= 5.828$$

$$SDR(humidity) = SD - SD(humidity)$$

$$= 8.7 - 5.828$$

$$= 2.872$$

windy

windy	mean	SD	CV	n	w(v)
false	32.66	6.80	20.85	3	3/5
True	39	12.72	32.5	2	2/5

$$SD(windy) = \frac{3}{5}(6.80) + \frac{2}{5}(12.72)$$

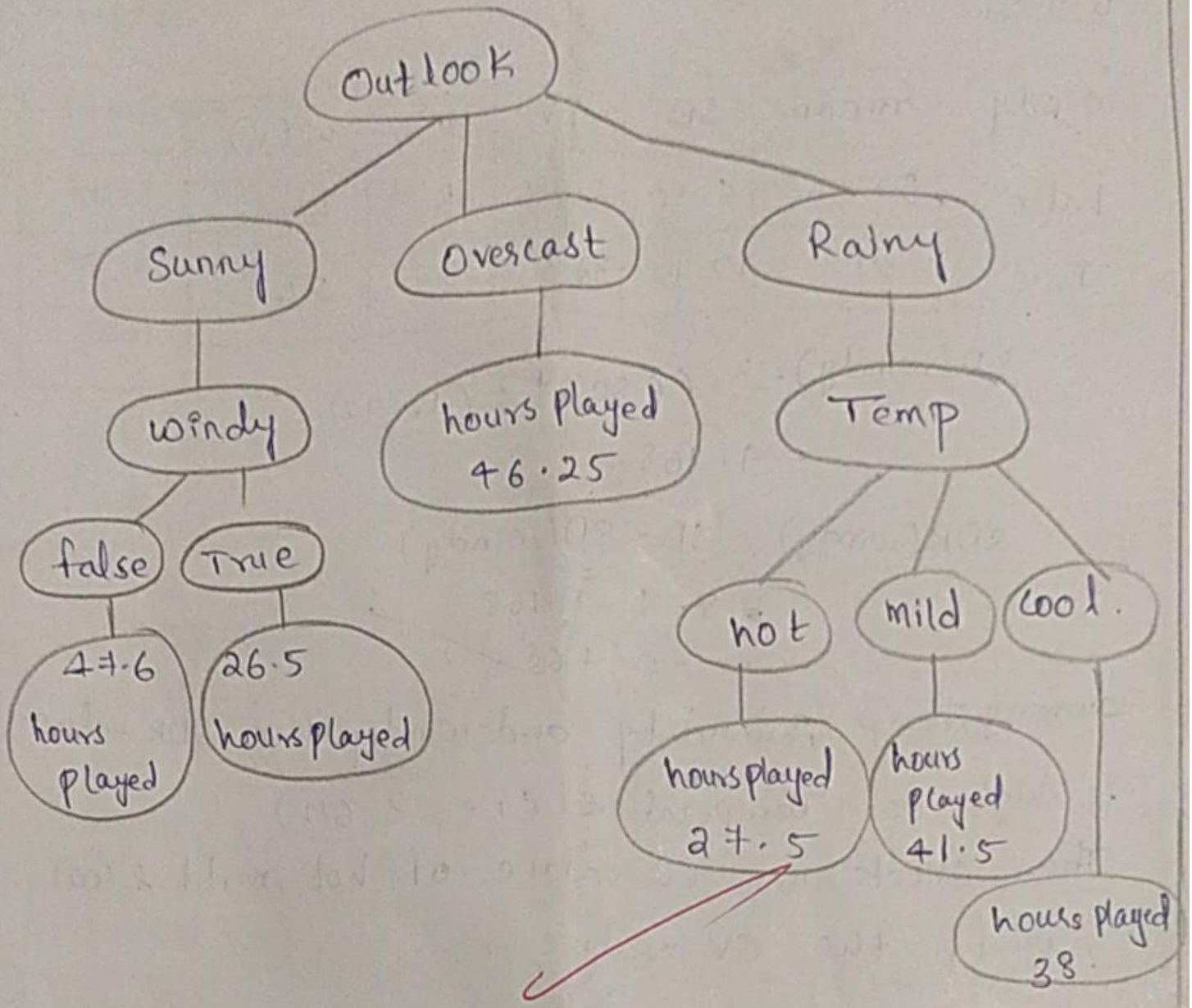
$$= 9.168$$

$$SDR(windy) = SD - SD(windy)$$

$$= 8.7 - 9.168$$

$$= -0.468$$

among, Temp, humidity and windy the SDR value is high for Temperature (i.e, 3.612)
Then, check for CV value of hot, mild & cool Satisfy the CV value.


Outlook
Sunny
Overcast
Rainy
Windy
hours Played
46.25
Temp
false
True
hot
mild
cool
47.6
hours
played
26.5
hours played
hours played
27.5
hours
Played
41.5
hours played
38

(10)

16/11/21