AI - Assignment - 7

19R41A05E7

(1)

Given data:

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
Overcast	cool	Normal	true	43
Rainy	mild	~~Normal~~ High	false	35
Rainy	cool	~~High~~ normal	false	38
Sunny	mild	Normal	false	46
Rainy	mild	Normal	~~false~~ true	48
Overcast	mild	~~Normal~~ high	true	52
Overcast	hot	Normal	false	44
Sunny	mild	High	true	30
Sunny	cool	Normal	true	23

Termination criteria: CV <= 10% (or)

minimum no. of samples.

(2)

Calculating mean, Standard deviation (SD), co-efficient of variation (CV)

$$\text{Mean} = \frac{\Sigma x}{n} = \frac{557}{14} = 39.78$$

$$SD = \sqrt{\frac{\Sigma (x - \text{mean})^2}{n}} = 9.67$$

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

Now, dataset is split into different attributes. The SD of each branch is calculated.

$$SD(\text{altr}) = \Sigma w(\text{branch}) \cdot SD(\text{branch})$$

and the result SDR (Standard Deviation Reduction) is calculated $SDR = SD - SD(\text{altr})$

$\therefore SD = 9.67$

Outlook:

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

$\therefore SD(\text{outlook}) = 5/4 * 8.7 + \frac{4}{14} * 4.03 + \frac{5}{14} * 12.2 = 8.59$

$SDR(\text{outlook}) = SD - SD(\text{outlook}) = 9.67 - 8.59 = 1.08$

Temperature:

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	8.38	19.65	6	6/14

SD(temp) = 4/14 * 10.34 + 4/14 * 12.14 + 6/14 * 8.36 = 10.01

SDR(temp) = SD − SD(temp) = 9.69 − 10.01 = −0.34

Humidity:

Humidity	Mean	SD	CV	n	w(v)
High	37.57	10.4	26.92	7	7/14 = 1/2
Normal	4.2	9.4	22.4	7	7/14 = 1/2

SD(humidity) = 7/14 * 10.11 + 7/14 * 9.4 = 9.77

SDR(humidity) = SD − SD(humidity)

= 9.67 − 9.77 = −0.1

Windy:

Windy	Mean	SD	CV	n	w(v)
True	37.6	11.6	30.8	6	6/14
False	41.3	8.41	20.3	8	8/14

SD(windy) = 6/14 * 11.6 + 8/14 * 8.41 = 9.77

SDR(windy) = SD − SD(windy) = 9.67 − 9.77 = −0.1

SDR
- windy → -0.1
- outlook → 1.08
- temp → -0.34
- humidity → -0.1
- windy → -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≤4)

Outlook

Overcast has CV of 8% which is less than threshold value.

Therefore, we need not go for further splitting

outlook
↓
overcast
↓
Hours played is 46.25

we need to split sunny & rainy ~~elements~~ columns

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

(5)

mean = 39.2, SD = 12.2 , CV = 31.0

Temperatures

Temp	Mean	SD	CV	n	w(v)
Mild	40.3	8.96	22.23	3	3/5
Cold	37.5	20.5	54.66	2	2/5

$SD(temp) = \frac{3}{5} * 8.96 + \frac{3}{5} * 20.5 = 13.576$

$SDR(temp) = SD - SD(temp) = 12.2 - 13.576 = -1.37$

Humidity:

Humidity	Mean	SD	CV	n	w(v)
High	37.5	10.6	28.26	2	2/5
Normal	40.3	15.3	37.96	3	3/5

$SD(humidity) = 2/5 * 10.6 + 3/5 * 15.3 = 13.46$

$SDR(Humidity) = SD - SD(humidity) = 12.2 - 13.42 = -1.22$

Windy:

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

$SD(windy) = 3/5 * 3.78 + 2/5 * 4.94 = 4.23$

$SDR(windy) = SD - SD(windy) = 12.2 - 4.23 = 7.97$

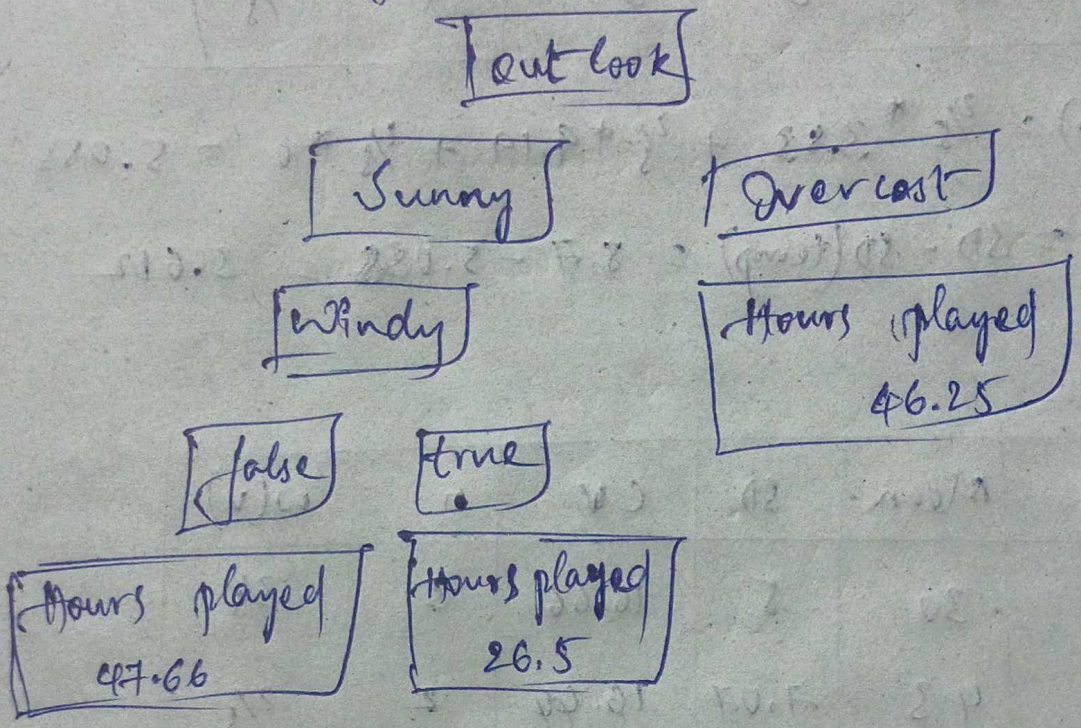
(6)

In outlook,

among 3 temp, humidity & windy ; SDR value is high for windy ; SDR = 7.97

Then, check for CV value

both true & false satisfy the CV value



Rainy:

Outlook	Temp	humidity	windy	hours played
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

(7)

Temperature:

Temp	Mean	SD	CV	n	w(v)
Hot	27.5	3.53	12.83	2	2/5
Mild	46.5	9.19	22.144	2	2/5
cool	38	0	0	1	1/5

SD(temp) = 2/5 * 3.53 + 2/5 * 9.19 + 1/5 * 0 = 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) = 3/5 * 5 + 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) = 3/5 * 6.80 + 2/5 * 12.78 = 9.168

SDR(windy) = SD - SD(windy) = 8.7 - 9.168 = -0.468

(8)

Among temp, humidity & windy, the SDR value is high for temperature (i.e. 3.612). Then check for CV value of hot, mild and cold satisfy the CV value.

- Design tree diagram to predict no. of hours to play based on weather conditions