15/11/2021 * ASSIGNMENT-7 * AI | 19K41A0594

Draw a decision tree diagram to predict number of hours to play based on whether conditions like outlook, temperature, humidity, windy. Consider dataset shown below.

Dataset: step1:

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 Standard deviation, CV, mean

$$\text{mean} = \frac{\Sigma x}{n} = \frac{25+30+46+45+52+23+43+35+38+46+48+52+44+30}{14}$$

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

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

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

Step3:

Dataset is splitted on different attributes, SD of each branch is calculated.

SD (attr) = $\Sigma w(\text{branch}) \cdot SD(\text{branch})$ and result is standard deviation reduction (SDR).

$$SDR = SD \cdot SD(attr)$$

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

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

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

$$= 8.59$$

$$SDR(\text{outlook}) = SD(\text{Target}) - SD(\text{outlook})$$

$$= 9.67 - 8.59$$

$$= 1.08$$

Temp:

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

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

$$= 10.01$$

$$SDR(\text{Temp}) = SD(\text{Target}) - SD(\text{Temp})$$

$$= 9.67 - 10.01$$

$$= -0.34$$

Humidity:

Humidity	Mean	SD	CV	n	w(v)
High	37.51	10.11	26.92	7	7/14
Normal	42	9.4	27.4	7	7/14

$$SD(Humidity) = \frac{7}{14} \times 10.11 + \frac{7}{14} \times 9.14$$
$$= 9.77$$

$$SDR(humidity) = SD(Target) - SD(humidity)$$
$$= 9.67 - 9.77$$
$$= -0.1$$

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

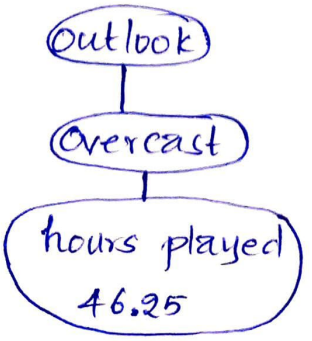
$$SD(windy) = \frac{6}{14} \times 11.6 + \frac{7}{14} \times 8.41$$
$$= 9.77$$

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

The value that has highest SDR is considered as root node

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:

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

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

$SDR(Temp) = 12.2 - 13.576$

$= -1.37$

Humid:

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

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

$= 13.42$

$SDR(Humid) = 12.2 - 13.42$

$= -1.22$

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

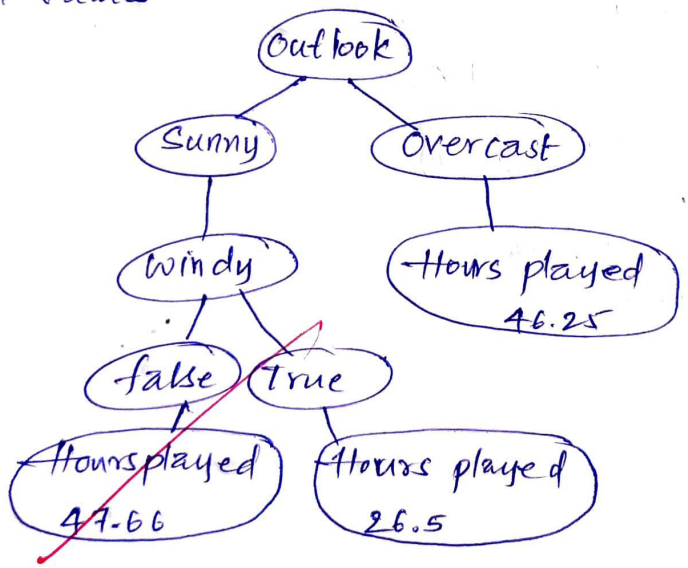
$$SD(windy) = \frac{3}{5}(3.78) + \frac{2}{5}(4.94)$$

$$= 4.23$$

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

⇒ In outlook, among temp, humidity and windy, SDR value is high in windy. SDR = 7.97

Then, check for CV value, both true and false satisfy 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	12.13	2	2/5
Mild	41.5	9.19	22.144	2	2/5
Cool	38	0	0	1	1/5

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

$= 5.088$

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

$= 8.7 - 5.088$

$= 3.612$

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, SDR value is high for temperature i.e 3.612

Then, check for CV value of hot, mild and cool satisfy CV value.