Asmath Fathima
19K41A04F3

AI - Assignment - 07

1) Draw a decision tree diagram to predict no. of hours to play based on weather conditions like outlook, temperature, humidity, windy. Consider dataset shown below.

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	Falle	45
sunny	cool	normal	False	52
overcast	cool	normal	True	43
rainy	Mild	high	False	35
rainy	cool	normal	False	38
sunny	mild	normal	True	46
rainy	mild	normal	True	48
overcast	mild	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 same

2) calculating mean, standard deviation (SD), co-efficient if variation (CV)

Asmath Fathima
19K41A04F3

AI - Assignment -07

1) Draw a decision tree diagram to predict no of hours to play based on weather conditions like outlook, temperature, humidity ,windy . Consider dataset shown below.

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	high	False	35
rainy	cool	normal	False	38
sunny	mild	normal	True	46
rainy	mild	normal	True	48
overcast	mild	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 same

2) Calculating mean, standard deviation (SD), co-efficient if 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, data set is split into diff. attributes. The SD of each branch is calculated.

$$SD(\text{attr}) = \Sigma\, W(\text{branch}) \cdot SD(\text{branch})$$

and the result SDR is calculated $SDR = SD - SD(\text{attr})$

$\therefore SD = 9.67.$

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	12.9	31.0	5	5/14

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

$$= 8.59$$

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

$$= 9.67 - 8.59 = 1.08$$

Temperature :

Temperature	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

3) SD (temperature) $= \frac{4}{14} * 10.34 + \frac{4}{14} * 12.14 + \frac{6}{14} * 8.38 - 10.01$

SDR (temp) = SD - SD (temp) = 9.67 - 10.01 = - 0.34

Humidity	mean	S.D	C.V.	n	W (HV)
High	37.51	10.11	26.92	7	7/14
normal	42	9.4	22.4	1	1/14

∴ SD (humidity) $= \frac{7}{14} \times 10.11 + \frac{7}{14} \times 9.4 = 9.77$

SDR (humidity) = SD - SD (humidity)

= 9.67 - 9.77 = -0.1

windy:

windy		mean	SD	CV	n	W(V)
True	8	37.6	11.6	30.8	6	6/14
False		41.3	8.41	20.3	8	8/14

∴ SD (windy) $= \frac{6}{14} * 11.6 + \frac{8}{14} * 8.41 = 9.77$

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

SDR (outlook) = 1.08

SDR (Temp) = -0.34

SDR (humidity) = -0.1

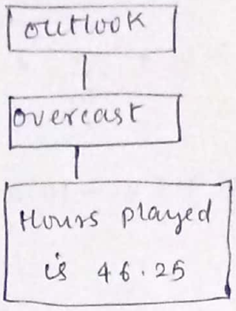
SDR (windy) = -0.1

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

considering termination criteria : CV is 10% / CV is ($n \leq 4$)

overcast has CV of 8% which is less than threshold.

4) value therefore, we need not go for further splitting.



we need to split sunny & rainy 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

∴ Mean = 39.2, SD = 12.2, CV = 31.0

Temperature :

Temperature	mean	SD	CV	n	w(v)
mild	40.3	8.96	22.33	3	3/15
cold	37.5	20.50	54.66	2	2/15

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

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

Humidity

Humidity	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

5. Windy :-

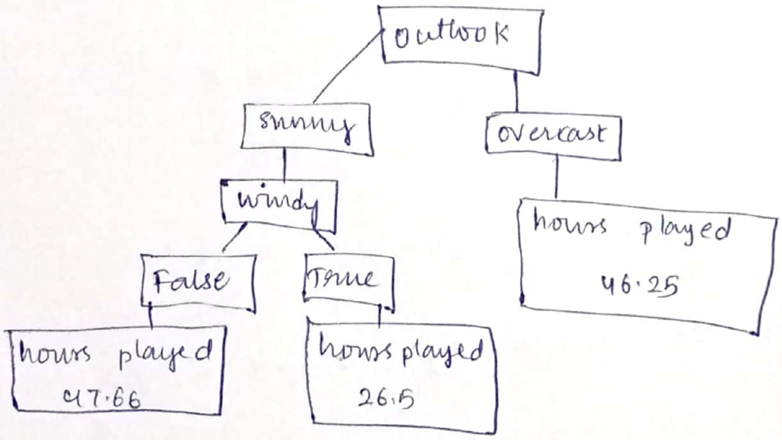
windy	mean	SD	cv	n	w(v)
False	47.66	3.78	7.94	3	3/5
True	26.5	4.94	18.65	2	2/5

$SD(windy) = \frac{3}{5} * 3.78 + \frac{2}{5} * 4.94 = 4.23$

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

In outlook, among temp, humidity & windy SDR value is high for windy SDR = 7.97.

Then, check for cv value both True & False satisfy the cvvalue.



Rainy :

outlook	Temp	humidity	windy.	hours played
Rainy	hot	high	False	25
Rainy	hot	high	True	30
Rainy	hot	high	False	35
Rainy	hot	normal	False	38
Rainy	hot	normal	True	48

6) Mean = 35.2, SD = 8.7, CV = 24.7

Temp:

Temp	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

$SD(temp) = \frac{2}{5} * 3.53 + \frac{2}{5} * 9.19 + \frac{1}{5} * 0 = 5.088$

$SDR(temp) = SD - SD(temp) = 8.7 - 5.088 = 3.612$

Humidity :

Humidity	mean	SD	CV	η	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 :

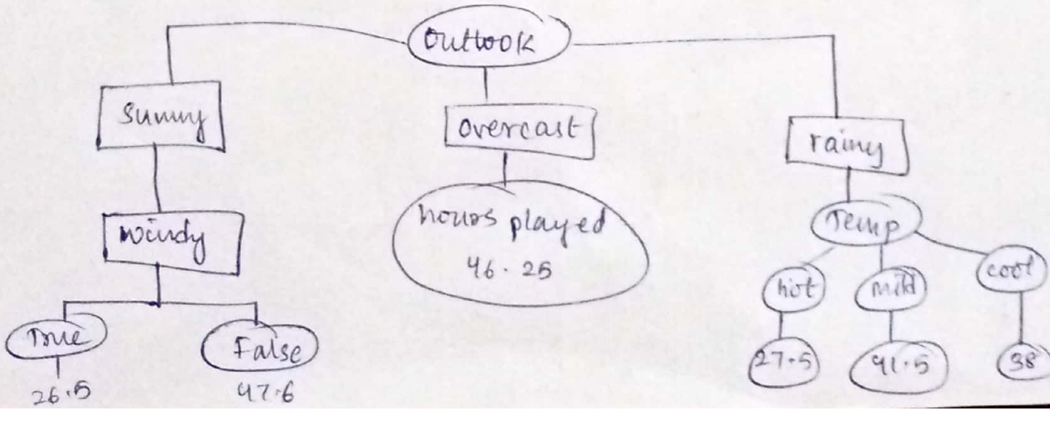
windy	mean	SD	CV	η	w(v)
False	32.66	6.80	20.85	3	3/5
True	39	12.72	32.5	2	2/5

$SD(windy) = \frac{3}{5} * 6.80 + \frac{2}{5} * 12.72 = 9.168$

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

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

7) Decision tree dig. to predict no. of hours of play based on weather conditions.


Outlook
Sunny
Overcast
rainy
Windy
hours played
46.25
Temp
True
False
26.5
47.6
hot
mild
cool
27.5
41.5
38