

AI- Assignment - 7

T. Phani prajna Goud
19K41A05B6

Draw a decision tree diagram to predict numbers 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	False	46
Rainy	mild	Normal	True	48
overcast	mild	High	False/True	52
overcast	Hot	Normal	False	44
Sunny	mild	High	True	30
Sunny	cool	Normal	True	23

Termination criteria : $CVL = 10\%$ on minimum number of samples = 4

Calculating mean, standard deviation (SD), coefficient of Variation (CV)

$$\text{mean} = \frac{\sum x}{n} = \frac{567}{14} = 39.78$$

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

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

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

$$SD(\text{attr}) = \sum w(\text{branch}) * SD(\text{branch})$$

And the result SDR (standard deviation reduction) 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	12.2	31.0	5	5/14

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

$$= 8.59$$

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

$$= 9.67 - 8.59 = 1.08$$

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

$$\therefore \text{SD (temperature)} = \frac{4}{14} \times 10.34 + \frac{4}{14} \times 12.14 + \frac{6}{14} \times 8.38$$

$$= 10.01$$

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

$$= 9.67 - 10.01 = -0.34$$

Humidity :-

Humidity	mean	SD	CV	n	w(Hu)
high	37.51	10.11	26.92	7	7/14
Normal	42	9.4	22.4	7	7/14

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

$$\begin{aligned} SDR(\text{humidity}) &= SD - SD(\text{humidity}) \\ &= 9.61 - 9.77 \\ &= -0.1 \end{aligned}$$

windy:

windy	mean	SD	LV	n	w(y)
True	37.6	11.6	20.8	6	6/14
False	41.3	8.01	20.3	8	8/14

$$\therefore SD(\text{windy}) = \frac{6}{14} \times 11.6 + \frac{8}{14} \times 8.01 = 9.77$$

$$\begin{aligned} SDR(\text{windy}) &= SD - SD(\text{windy}) \\ &= 9.61 - 9.77 = -0.1 \end{aligned}$$

$$SDR(\text{out look}) = 1.08$$

$$SDR(\text{Temperature}) = -0.34$$

$$SDR(\text{Humidity}) = -0.1$$

$$SDR(\text{windy}) = -0.1$$

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

$$\text{mean} = 34.2$$

$$SD = 12.2$$

$$CV = 31.0$$

Temperature

Temperature	mean	SD	CV	n	w(v)
mild	40.3	8.96	22.33	3	3/5
cold	37.5	20.50	54.66	2	2/5

$$SD(\text{temperature}) = \frac{3}{5} \times 8.96 + \frac{2}{5} \times 20.50$$

$$= 13.576$$

$$SDR(\text{temperature}) = SD - SD(\text{temp})$$

$$= 12.2 - 13.576 = -1.37$$

Humidity:

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

$$SD(\text{humidity}) = \frac{2}{5} \times 10.6 + \frac{3}{5} \times 15.30$$

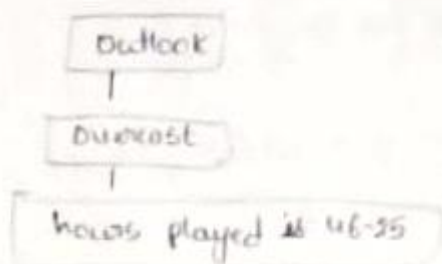
$$= 13.42$$

considering termination criteria

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

outlook

overcost has CV of 8% which is less than threshold value therefore, we need not go for further splitting



we need to split sunny and rainy columns:

Sunny:

outlook	Temperature	Humidity	Windy	hours played
Sunny	mild	high	False	46
Sunny	cool	normal	False	52
Sunny	cool	normal	True	23
Sunny	mild	normal	False	46
Sunny	mild	high	True	30

$$\begin{aligned} \text{SDR}(\text{Humidity}) &= \text{SD} - \text{SD}(\text{humidity}) \\ &= 12.2 - 13.42 \\ &= -1.22 \end{aligned}$$

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

$$\begin{aligned} \text{SD}(\text{windy}) &= 3/5 \times 3.78 + 2/5 \times 4.94 \\ &= 4.23 \end{aligned}$$

$$\begin{aligned} \text{SDR}(\text{windy}) &= \text{SD} - \text{SD}(\text{windy}) \\ &= 12.2 - 4.23 = 7.97 \end{aligned}$$

In outlook

among. Temperature, humidity and windy SPR value is high for windy $\text{SDR} = 7.97$

both True and false satisfy CV value

Temperature

Temp	mean	SD	CV	n	w(CV)
Hot	27.5	3.53	12.83	2	2/5
mild	41.5	9.19	22.144	2	2/5
Cold	38	0	0	1	1/5

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

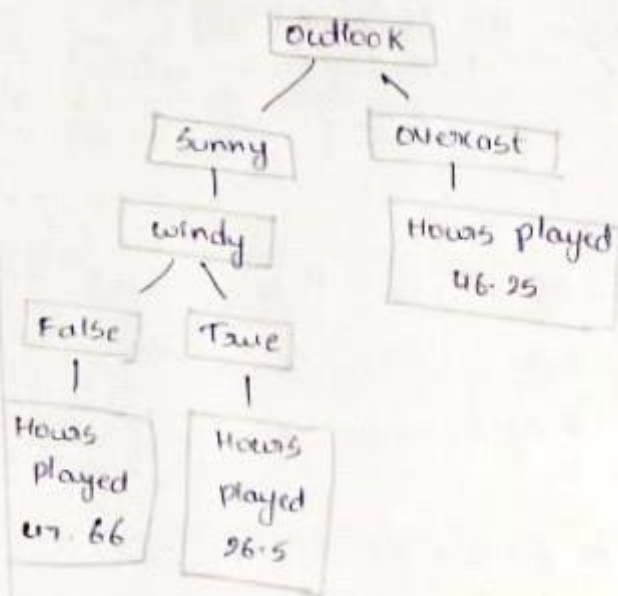
$$SDR(\text{temperature}) = SD - SD(\text{temp}) \\ = 8.7 - 5.088 \\ = 3.612$$

Humidity

Humidity	mean	SD	CV	n	w(CV)
High	30	5	16.66	3	3/5
Normal	43	7.07	16.44	2	2/5

$$SD(\text{humidity}) = \frac{3}{5} \times 5 + \frac{2}{5} \times 7.07 = 5.828$$

$$SDR(\text{humidity}) = SD - SD(\text{humidity}) \\ = 8.7 - 5.828 = 2.872$$



Rainy:

outlook	Temperature	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

windy	mean				
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

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

$$= 9.168$$

$$SPR(windy) = SD - SD(windy)$$

$$= 8.7 - 9.168$$

$$= -0.468$$

Among Temperature, humidity and windy

The SPR value is high for temperature (i.e., 3.692)

The check for cv value of hot, mild and cold satisfy the cv value.

Design tree diagram to predict number of hours to play based on weather conditions

