

AI ASSIGNMENT

(1)

Name : G. Rama Krishna

Branch : EEE-3B

Pin : 20K45A0240

Draw a decision tree diagram to predict number 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	Cool	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	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30
Sunny	Cool	Normal	True	23

Termination Criteria : $CV \leq 10\%$ or minimum number of samples.

Calculating mean, standard deviation (SD), Co-efficient of variation (CV)

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

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

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

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

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

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

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

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

$$= 9.67 - 8.59$$

$$= 1.08$$

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

$$\therefore SD(\text{temperature}) = \frac{4}{14} \times 10.34 + \frac{4}{14} \times 12.14 + \frac{6}{14} \times 8.38 = 10.01$$

$$SDR(\text{temperature}) = SD - SD(\text{temperature}) = 9.69 - 10.01 = -0.34$$

Humidity

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

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

$$SDR(\text{Humidity}) = SD - SD(\text{humidity})$$

$$= 9.67 - 9.77 = -0.1$$

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

$$\therefore SD(\text{windy}) = \frac{6}{14} \times 11.6 + \frac{8}{14} \times 8.41 = 9.77$$

$$SDR(\text{windy}) = SD - SD(\text{windy}) = 9.67 - 9.77 = -0.1$$

$$\text{Hence, } SDR(\text{outlook}) = 1.08$$

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

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

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

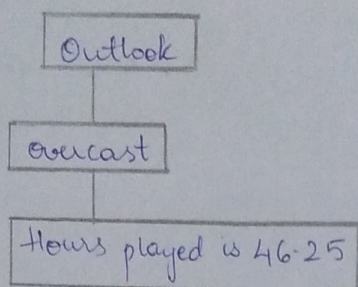
The value that has highest SDR is considered as root node (i.e. decision node)

Considering termination criteria

CV is 10% of CV is ($n \leq 1$)

outlook

overcast has CV of 8% which is less than threshold value. Therefore, we need not to go for further splitting



we need to split sunny & rainy columns

Outlook	Temperature	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

\therefore Mean = 39.2, SD = 12.2 ; CV = 31.0

Temperature:

Temperature	Mean	SD	CV	n	w(V)
Mild	40.3	8.96	22.23	3	3/5
Cold	37.5	20.50	54.66	2	2/5

$$SD(\text{Temperature}) = \frac{3}{5} \times 8.96 + \frac{2}{5} \times 20.5 = 13.576$$

$$SDR(\text{Temperature}) = SD - SD(\text{Temperature}) = 12.2 - 13.576 = -1.37$$

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

$$SD(\text{humidity}) = \frac{2}{5} \times 10.6 + \frac{3}{5} \times 15.3 = 13.44$$

$$SDR(\text{humidity}) = SD - SD(\text{humidity}) = 12.2 - 13.44 = -1.22$$

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

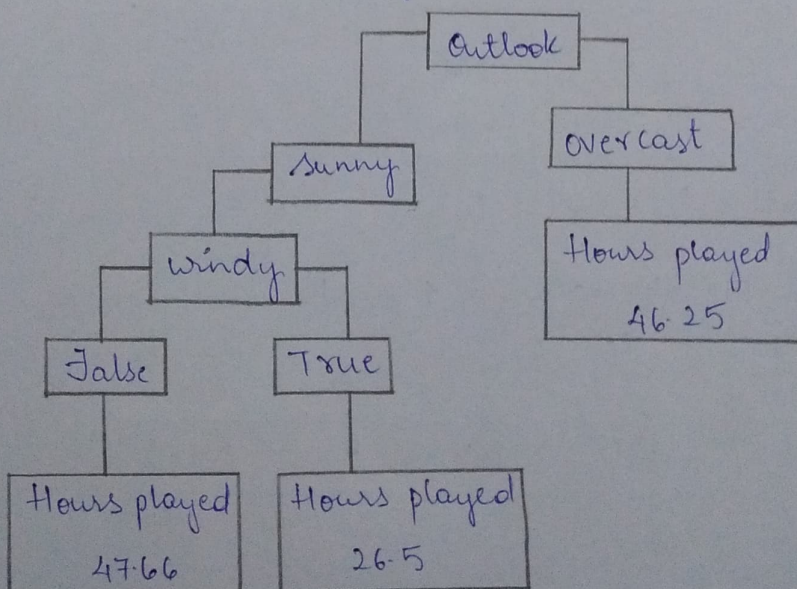
$$SD(\text{windy}) = \frac{3}{5} \times 3.78 + \frac{2}{5} \times 4.94 = 4.23$$

$$SDR(\text{windy}) = SD - SD(\text{windy}) = 12.2 - 4.23 = 7.97$$

In outlook; among temperature, 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 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

$$\text{Mean} = 35.2, \text{SD} = 8.7, \text{CV} = 24.7$$

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

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

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

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

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

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

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

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

$$SDR(\text{windy}) = SD - SD(\text{windy}) = 8.7 - 9.168 = -0.468$$

Among; temperature, 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.

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

