AI - ASSIGNMENT - 7

19k41A0536
DHARMULA. AKHIL

Draw a decision tree diagram to predict numbers & hours to play board or whether conditions like outlook, temperature, humidity, windy. Consider data set below (pdf) Termination criteria: CVC=10% of minimum number of sample :4.

Sol Calculating mean, standard deviation (SD), co-efficient & 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 we split into different attributes. The SD & each branch is Calculated.

$$SD(attr) = \Sigma w(branch) . SD(branch).$$

and the result SDR (Standard deviation reduction) is Calculated $SDR = SD - SD(attr)$

$\therefore SD = 9.67$

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	81.0	5	5/14.

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

$$= 8.59$$

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

$$= 1.08$$

536 (10)

Temperature:

Temperature	mean	SD	CV	n	w(v)
Hot	36.25	10.34	30.6	4	4/14
Cold	Mild 39	12.14	31.1	4	4/14
Mild	42.6	8.38	19.65	6	6/14

$\therefore SD(temp) = 4/14 * 10.34 + 4/14 * 12.14 + 6/14 * 8.38$

$= 10.01$

$SDR(temp) = SD - SD(temp)$

$= 9.67 - 10.01$

$= -0.34$.

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.

$\therefore SD(Hu) = 7/14 * 10.11 + 7/14 * 9.4 = 9.77$

$SDR(Hu) = SD - SD(Hu)$

$= 9.67 - 9.77 = -0.1$

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.

$\therefore SD(windy) = 6/14 * 11.69 + 8/14 * 8.41 = 9.77.$

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

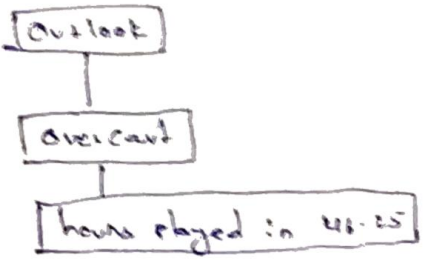
The value that has highest SDR is considered as root node (i.e decision node).

Considering termination criteria.

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

Outlook

...cart has cv & cv which is 19K41A0536 ①
... than threeshold value therefore, we need not go to further splitting.



we need to split sunny and rainy Columns.

Sunny:

outlook	temperature	Humidity	windy	hours played
Sunny	mild	high	False	47
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	22.23	3	~~3/5~~ 3/5
Cold	35.5	20.50	54.66	2	2/5

SD(temp) = 3/5 * 8.96 + 2/5 * 20.50

= 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	15.30	37.96	3	3/5

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

$= 13.42$

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

$= 12.2 - 13.42$

$= -1.22$

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

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

$= 4.23$

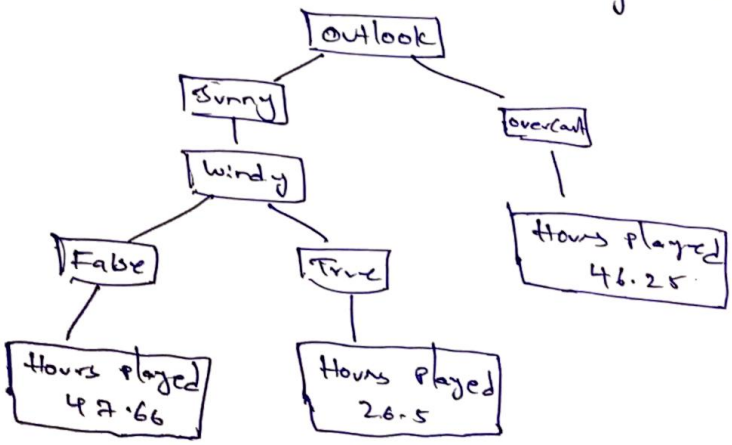
$SDR(\text{windy}) = SD - SD(\text{windy})$

$= 12.2 - 4.23 = 7.97$

In outlook, among temperature, humidity and windy SDR value is high for windy SDR = 7.97.

Then check for Cv value.

both true and false statisfy 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

mean = 35.2 , SD = 8.7 , CV = 24.7

Temperature

Temp	mean	SD	CV	n	w(x)
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

$SD(temp) = 2/5 \times 3.53 + 2/5 \times 9.19 + 1/5 \times 0$

$= 5.088$

$SDR(temp) = SD - SD(temp)$

$= 8.7 - 5.088$

$= 3.612.$

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

$SD(Hu) = 3/5 \times 5 + 2/5 \times 7.07 = 5.828$

$SDR(humidity) = SD - SD(Hu)$

$= 8.7 - 5.828$

$= 2.872$

windy

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

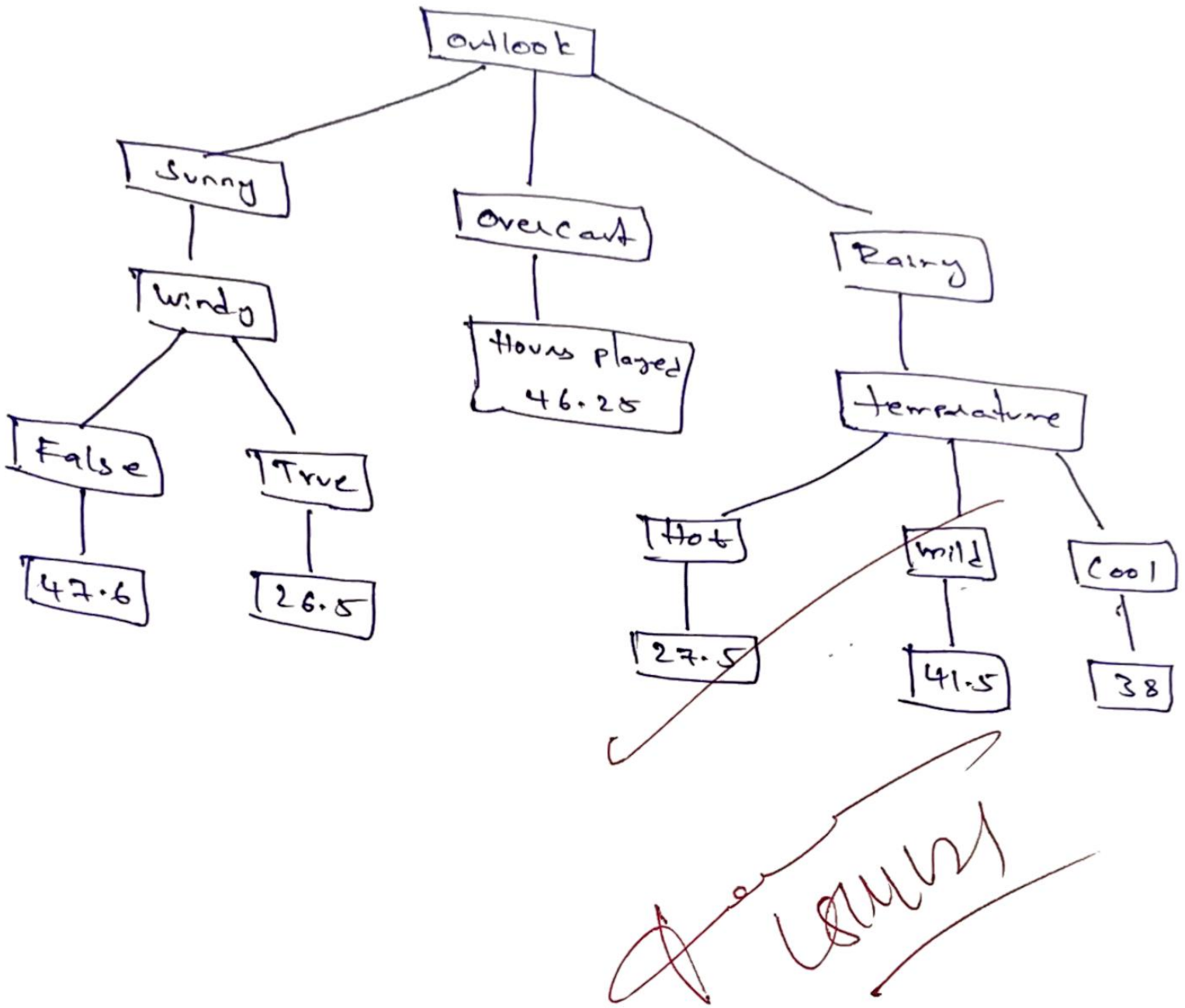
$SD(windy) = 3/5 \times 6.80 + 2/5 \times 12.72$

$= 9.168$

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

$= -0.468.$

Among temperature, humidy and windy the SDR valve is high for temperature i.e (3.612) the check for the CV valve of hot, mild, cold. satisfy the CV valve.

19K41A0536 ⑦

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