19K41A04A6
D. Meghana.

Draw a decision tree diagram to predict number of hours to play based on weather conditions like outlook, temperature, humidity, windy.

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

Calculating mean, Standard deviation (SD), coefficient of 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 different attributes
The SD of each branch is calculated

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

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

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

$= 8.59$

$\therefore$ SDR (outlook) = SD − 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

$\therefore$ SD (Temperature) = 4/14 × 10.34 + 4/14 × 12.14 + 6/14 × 8.38

= 10.01

SDR(Temperature) = SD − SD(Temperature)

= 9.67 − 10.01

= −0.34

Humidity:

humidity	mean	SD	CV	n	w(H.V)
high	37.51	10.11	26.92	7	7/14
normal	42	9.4	22.4	7	7/14

$\therefore$ SD (humidity) = 7/14 × 10.11 + 7/14 × 9.4

= 9.77

SDR(humidity) = SD − 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

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

$= 9.77$

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

$= 9.67 - 9.77$

$= -0.1$

$\therefore SDR(outlook) = 1.08$

$SDR(Temperature) = -0.34$

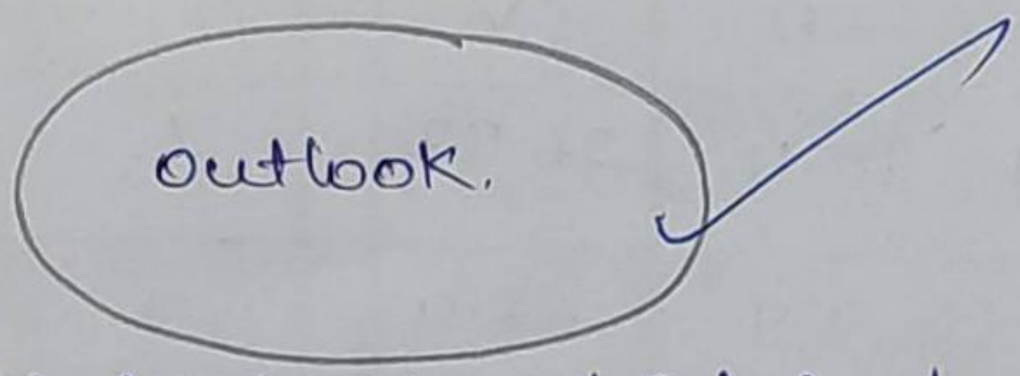
$SDR(humidity) = -0.1$

$SDR(windy) = -0.1$

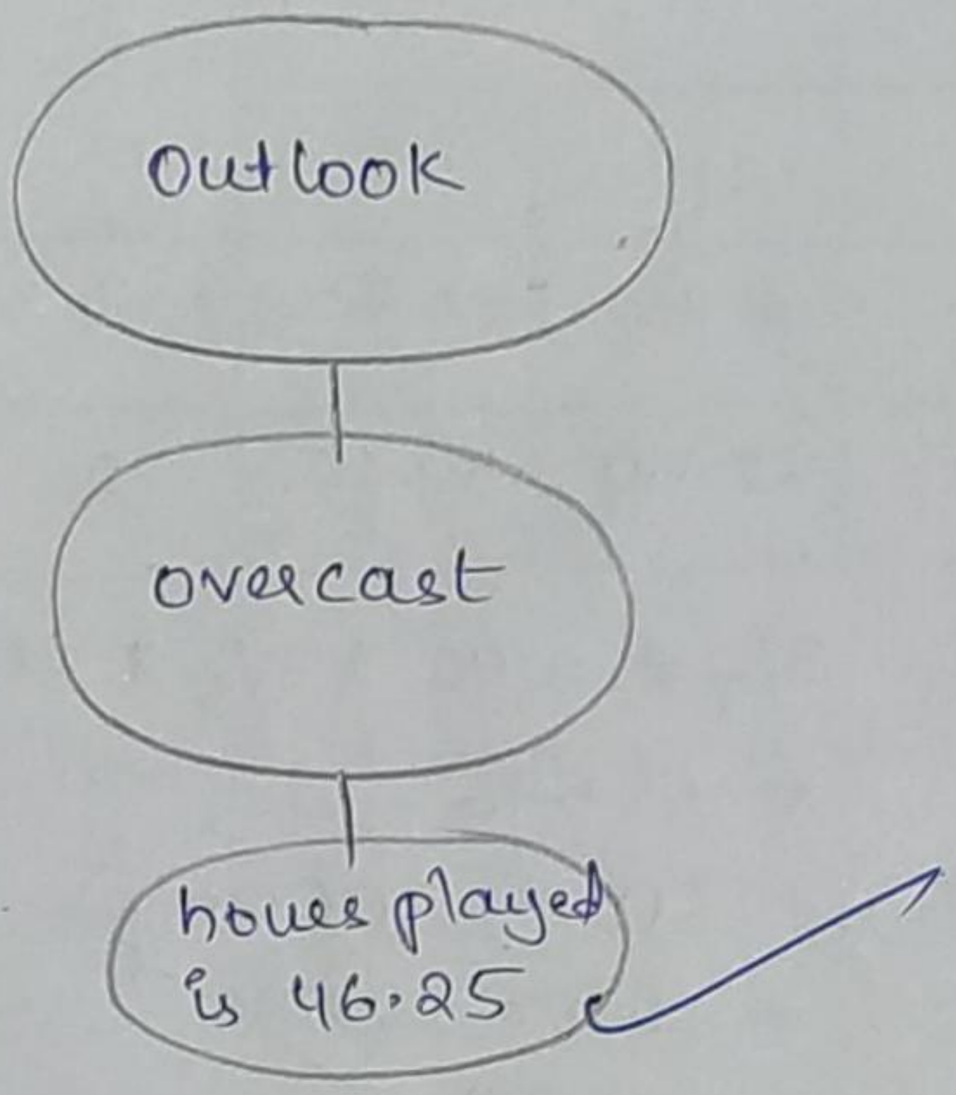
The value that has highest SDR is considered as root node (i.e decision node)

Considering termination criteria

CV is 10%. or CV is (n ≤ 4)



Overcast has CV of 8%. which is less than threshold value therefore we need not to further split



We need to split Sunny & Rainy columns.

Sunny:

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

∴ mean = 39.2

∴ SD = 12.2

∴ CV = 31.0

Temperature:

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(Temperature) = 3/5 * 8.96 + 2/5 * 20.50

= 13.576

SDR(Temperature) = 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(humidity) = 2/5 * 10.6 + 3/5 * 15.30

= 13.42

SDR(humidity) = SD − SD(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

$SD(windy) = 3/5 * 3.78 + 2/5 * 4.94$

$= 4.23$

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

$= 12.2 - 4.23$

$= 7.97$

In outlook

among, Temp, humdity and windy

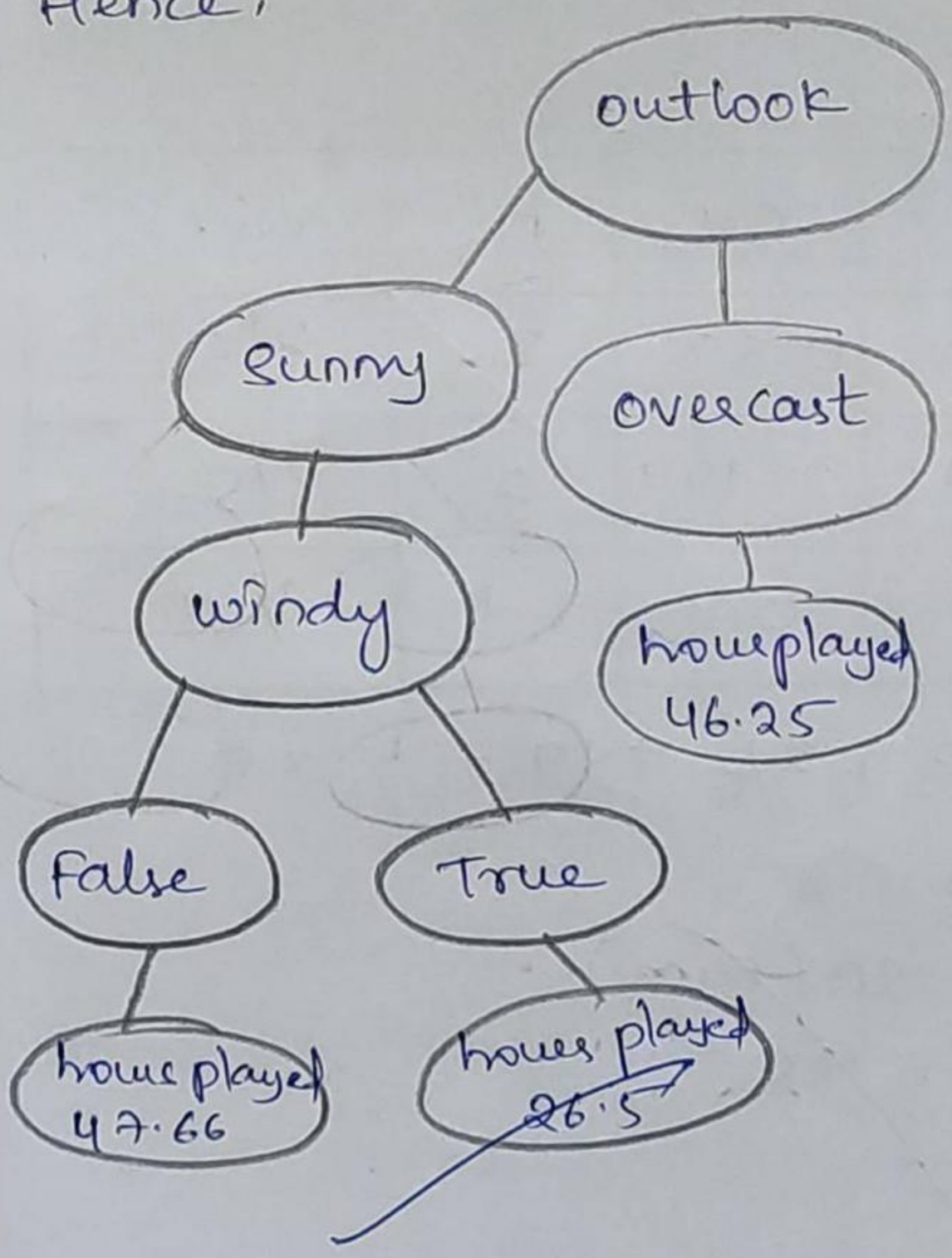
SDR value is high for windy

$\therefore$ SDR = 7.97.

Then, check for CV value.

both True, & false satisfy the CV value. (i.e $n \leq 4$)

Hence,



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:

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
Cool	38	0	0	1	1/5

SD(Temp) = 2/5 * 3.53 + 2/5 9.19 + 1/5 x 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) = 3/5 × 5 + 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) = 3/5 × 6.80 + 2/5 × 12.72

= 9.168

SDR(windy) = S(D) - SD(windy)

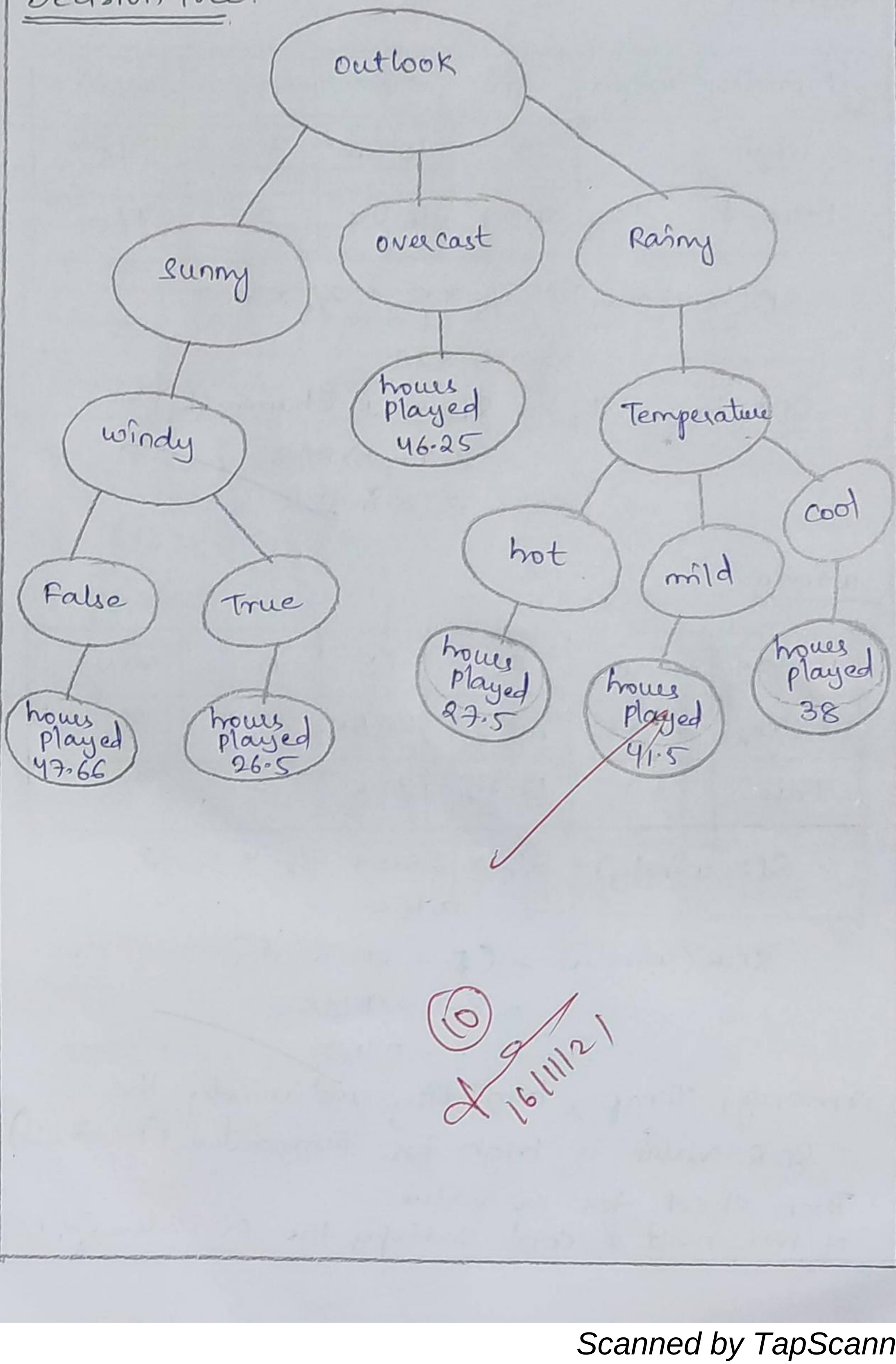
= 8.7 - 9.168

= -0.468

among, Temp, humidity and windy the SDR value is high for Temperature (i.e 3.612)

Then check for CV value of hot, mild & cool satisfy the CV value (i.e. n≤4)

Decision Tree:



(10)
16/11/21