STEP-1

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

STEP-2

Calculate SD, CN, mean

$$\text{mean} = \frac{\Sigma x}{n}$$

$$\text{mean} = \frac{25+30+46+45+52+23+43+35+38+46+48+52+44+30}{14}$$

$$= \frac{557}{14} = 39.78$$

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

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

STEP-3

Dataset is split on different attributes the SD of each branch is calculated

$$SD(attr) = \sum w(branch)\, SD(branch)$$

$$SDR = SD - SD(attr)$$

$\therefore SD(Target) = 9.67$

Outlook

	mean	SD	CV	n	w(r)
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

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

$$SDR(outlook) = SD(Target) - SD(outlook)$$

$$= 9.67 - 8.59 = \underline{1.08}.$$

Temp:

	mean	SD	CV	n	w(r)
hot	36.25	10.34	30.6	4	4/14
cool	39	12.14	31.1	4	4/14
mild	42.6	3.38	19.65	6	6/14

$$SD(Temp) = \frac{4}{14}(10.34) + \frac{4}{12}(12.14) + \frac{6}{14}(3.38) = 10.01$$

$$SDR(Temp) = 9.67 - 10.01 = -0.34$$

Humidity:

	mean	SD	CV	n	w(r)
high	39.51	10.11	26.92	7	7/14
normal	42	9.4	27.4	7	7/14

SD(humidity) = $\frac{7}{14} \times 10.11 + \frac{7}{14} \times 9.14 = 9.77$

SDR(humidity) = 9.69 − 9.77 = −0.1

windy :

	mean	SD	cv	n	w(r)
True	39.6	11.6	30.8	6	6/14
False	41.3	8.44	20.3	8	8/14

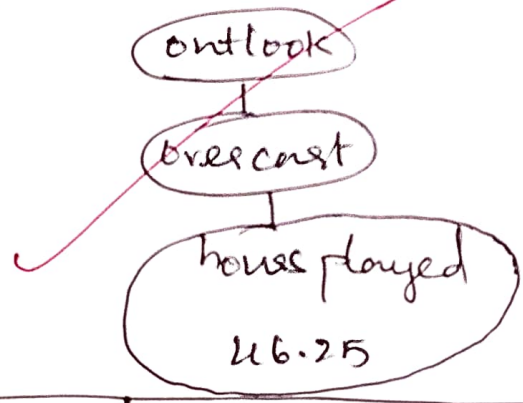
SD(windy) = $\frac{6}{14} \times 11.6 + \frac{7}{14} \times 8.41 = 9.77$

SDR(windy) = 9.69 − 9.77 = −0.1

the value that has highest SDR is considered as root node.

Considering termination Criteria cv is 10% & n≤4

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



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

Temp :

	mean	SD	CV	n	w(r)
mild	40.3	8.96	22.23	3	3/5
cool	39.5	20.50	54.66	2	2/5

$$SD(Temp) = \frac{3}{5}(8.96) + \frac{2}{5}(20.50) = 13.596$$

$$= 12.2 - 13.596 = -1.37$$

Humidity

	mean	SD	CV	n	w(r)
high	39.5	10.6	2826	2	2/5
normal	40.3	15.30	37.96	3	3/5

$$SD(humid) = \frac{2}{5}(10.6) + \frac{3}{5}(15.30) = 13.42$$

$$SD = 12.2 - 13.42 = -1.22$$

windy

	mean	SD	CV	n	w(r)
false	47.66	3.78	7.94	3	3/5
True	26.5	4.98	18.65	2	2/5

$$SD(windy) = 3/5\,(3.78) + 2/5\,(4.94) = 4.23$$

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

then check for highest SDR.

In outlook, among temperature, humidity & windy SDR value is high for windy (SDR = 7.97)

then, check for CV value.

Both True & false satisfy the CV value.