# Assignment - 7

19K41A0538
CSE - Sec - A

Draw a decision tree diagram to Predict number of hours to play based on weather conditions like outlooks, 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 | True | 57 |
| overcast | Hot | normal | false | 44 |
| Sunny | mild | high | true | 30 |
| Sunny | cold | Normal | True | 23 |

Termination criteria: CV < = 10% or Minimum number of samples: 4

538 (10)

2/11/21

calculating mean, standard deviation (SD), co-effiient of Variation (CV)

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

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

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

Now, dataset is split into diffrent attributes. The SD of each branch is calculated

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

and the result SDR (standard deviation reduction) is calculated

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

$$\therefore SD = 7.67$$

outlook :-

| outlook | Mean | SD | CV | n | w(v) |
|---|---|---|---|---|---|
| Rainy | 35.2 | 8.7 | 24.7 | 5 | 5/14 |
| overcast | 46.~~5~~25 | 4.03 | 8.72 | 4 | 4/14 |
| Sunny | 39.2 | 12.2 | 81.0 | 5 | 5/14 |

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

$$SDR(\text{outlook}) = SD - SD(\text{outlook}) = 9.67 - 8.59 = 1.08$$

Temperature :-

| Temperature | Mean | SD | CV | n | w(v) |
|---|---|---|---|---|---|
| hot | 36.25 | 10.34 | 3.06 | 4 | 4/14 |
| cold | 39 | 12.14 | 31.1 | 4 | 4/14 |
| mild | 42.6 | 8.38 | 19.65 | 6 | 6/14 |

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(v) |
|---|---|---|---|---|---|
| High | 37.51 | 10.11 | 26.92 | 7 | 7/14 |
| Normal | 42 | 9.4 | 22.4 | 7 | 7/14 |

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

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 |

∴ SD (windy) = 6/14 * 11.6 + 8/14 * 8.41 = 9.77

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

SDR (outlook) = 1.08

SDR (Temperature) = -0.34

SDR (Humidity) = -0.1

SDR (windy) = -0.1

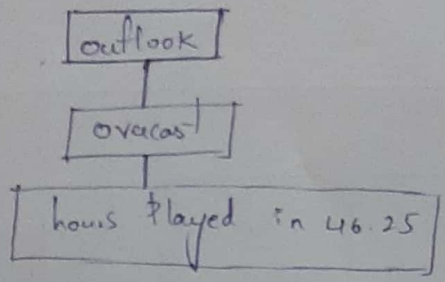
The value that has highest SDR is considered as root node (i.e, decision node)

Considering termination criteria

CV is 10% or CV is (n<4)

outlook is

overcast has cv of c.l 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 | 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 :-

| Temperature | mean | SD | CV | n | w(v) |
|---|---|---|---|---|---|
| mild | 40.3 | 8.96 | 22.23 | 3 | 3/5 |
| cold | 35.5 | 20.50 | 54.66 | 2 | 2/5 |

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

= 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) = $\frac{2}{+5} \times 10.6 + \frac{3}{5} \times 15.36$

= 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 |

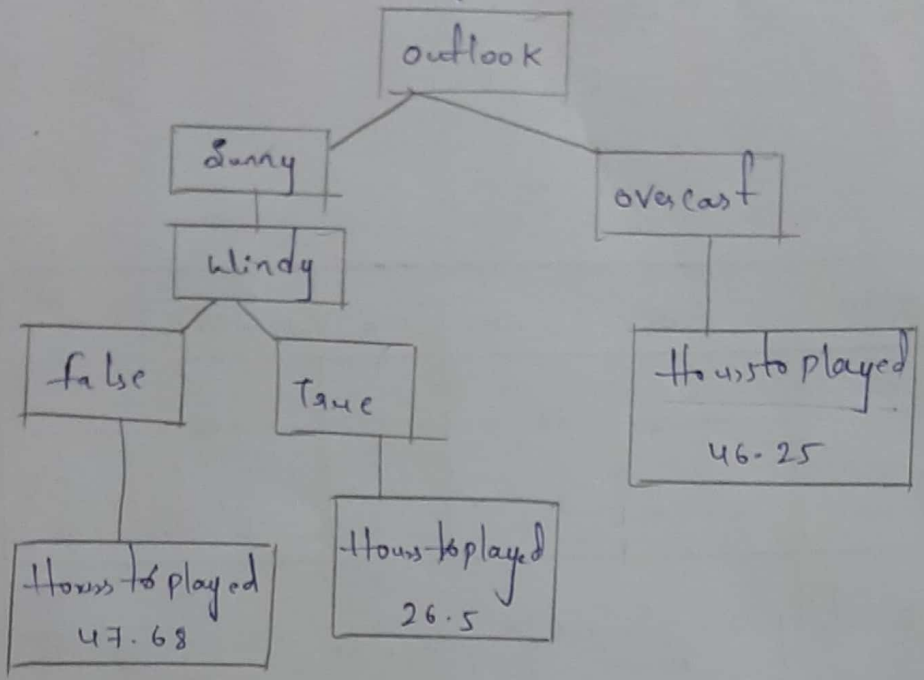
SD (windy) = $\frac{3}{5} \times 3.78 + 2/5 \times 4.94$

= 4.23

SDR (windy) = SD - SD (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 satisfy the CV value




Rainy :-

| outlook | Temperature | humidity | windy | hours to played |
|---|---|---|---|---|
| Rainy | hot | high | false | 25 |
| Rainy | hot | high | True | 30 |
| Rainy | mild | high | false | 35 |
| Rainy | cold | normal | false | 38 |
| Rainy | mild | normal | true | 48 |

Mean = 35.2, SD = 8.7, 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 |

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

$$= 5.088$$

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

$$= 8.7 - 5.088$$

$$= 3.612$$

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 |

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

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

$$= 8.7 - 5.828$$

$$= 2.872$$

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 |

$$SD(windy) = \frac{3}{5} \times 6.80 + \frac{2}{5} \times 12.72$$
$$= 9.168$$
$$SDR(windy) = SD - SD(windy)$$
$$= 8.7 - 9.168$$
$$= -0.468$$

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

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