A1 - Assignment - 7

19K41A05E4

Sriharsha.s

Given data:

Outlook	Temp	Humidity	windy	Hourtrophy
Rainy	Hot	High	False	25
Rainy	Hot	high	True	30
Overcat	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	52
overcast	hot	Normal	False	44
Sunny	mild	High	True	30
Sunny	cool	Normal	True	23

Termination criteria: cvc = 10% (∞)

minimum no. of samples.

Calculating mean, standard deviation (SD), co-efficient of variation (cv).

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

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

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

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

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

and the result, SDR (standard Deviation Reduction) is
calle calculated, $SDR = SD - SD(\text{after})$

$$\therefore SD = 9.67$$

outlook:

Outlook	Mean	SD	CV	n	W(v)
Raining	35.2	8.7	24.7	5	5/14
Overcast	46.25	4.03	6.72	4	4/14
Sunny	39.2	12.2	81.0	5	5/14

$$\therefore SD(\text{outlook}) = \frac{5}{14} * 8.7 + \frac{4}{14} * 4.03 + \frac{5}{14} * 12.2 = 8$$

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

Temperature:

Temp	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

$$SD(temp) = \frac{4}{14} * 10.34 + \frac{4}{14} * 12.14 + \frac{6}{14} * 8.36 = 10.01$$

$$SDR(temp) = SD - SD(temp) = 9.69 - 10.01 = -0.34$$

Humidity:

Humidity	Mean	SD	CV	n	W(v)
High	37.51	10.11	26.92	7	7/14 = 1/2
Normal	4.2	9.4	22.4	7	7/14 = 1/2

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

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

$$9.67 - 9.77 = -0.1$$

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

$SD \text{ (windy)} = \frac{6}{14} * 11.6 + \frac{8}{4} * 8.41 = 9.77$

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

$$SDR \begin{cases} \text{windy} & \rightarrow -0.1 \\ \text{Outlook} & \rightarrow 1.08 \\ \text{Temp} & \rightarrow -0.34 \\ \text{Humidity} & \rightarrow -0.1 \\ \text{windy} & \rightarrow 0.1 \end{cases}$$

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

considering termination criteria

cv is 10% (or) cv is $(n \leq 4)$.

Over cast has cv of 8% which is less than threshold value.

∴, we need not go for further splitting

```
   ┌─────────┐
   │ outlook │
   └─────────┘
        │
        ↓
   ┌──────────┐
   │ overcast │
   └──────────┘
        │
        ↓
┌──────────────────────────┐
│ Hours played is  46.25   │
└──────────────────────────┘
```

Outlook	Temp	Humidity	Windy	Hours played
Sunny	mild	High	False	45
Sunny	mcool	Normal	False	52
Sunny	Cool	Normal	True	23
Sunny	Grotd	Normal	True	46
Sunny	mild	High	True	30
Sunny				

mean = 39.2   SD = 12.2   CV = 31.0

Temperature:

Temp	Mean	SD	CV	n	N(v)
mild	40.3	8.96	22.23	3	$3/5$
cold	37.5	20.5	54.66	2	$2/5$

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

$$SDR(humidity) = SD - SD(humidity) = 12.2 - 13.42$$

$$= -1.22$$

Windy:

Windy	Mean	SD	CV	n	N(v)
False	47.66	3.78	7.94	3	$3/5$
True	26.5	4.94	18.63	2	$2/5$

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

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

In outlook:

Among, temp, humidity & windy, SDR value is high for

windy: $SDR = 7.97$.

Then, check for cv value.

both true & false satisfy the cv value.

```
                    ┌──────────┐
                    │ outlook  │
                    └──────────┘
         ┌──────────┐              ┌──────────┐
         │ Sunny    │              │ overcast │
         └──────────┘              └──────────┘
              ┌─────────┐          ┌─────────────┐
              │ windy   │          │ Hours played│
              └─────────┘          │   46.25     │
         ┌───────┐ ┌──────┐        └─────────────┘
         │ False │ │ true │
         └───────┘ └──────┘
   ┌──────────────┐  ┌──────────────┐
   │ Hours played │  │ Hours played │
   │   47.66      │  │    26.5      │
   └──────────────┘  └──────────────┘
```

Rainy:

outlook	Temp	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(r)
Hot	27.5	3.53	12.13	2	2/5
mild	41.5	9.19	22.144	2	2/5
cool	38	0	0	1	1/5

$SD(temp) = \frac{2}{5} \times 3.53 + \frac{2}{5} \times 9.19 + \frac{1}{5} \times 0 = 5.068$

$SDR(temp) = SD - SD(temp) = 8.7 - 5.068 = 3.612$

Humidity :

Humidity	Mean	SD	CV	n	w(v)
high	36	5	16.66	3	3/5
normal	43	7.07	16.44	2	2/5

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

$SDR(humidity) = SD - SD(humidity) = 8.7 - 5.828$

$= 2.872$

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

$$SD(windy) = 3/5 \approx 6.80 + 2/8 = 12.78 = 9.168$$

$$SDR(windy) = SD - SD(windy) = 8.7 - 9.168 = -0.468$$

Ⓢ

Among temp, 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.

- Design tree diagram to predict no of hours of play
based on weather conditions.

Scanned with CamScanner