

ASSIGNMENT - 7

19K41A0555

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

Termination criteria:  $cv \leq 10\%$  or minimum number of samples = 4

Root node identification :-

Step 1 :- Standard deviation of play hours =  $SD(\text{hours}) = 9.3210$

Step 2 :- finding the standard deviation reduction of the attributes

Outlook	Mean	SD	Count	$w(O)$
Sunny	39.2	10.8701	5	5/14
Rainy	35.2	7.7820	5	5/14
Overcast	46.25	3.4910	4	4/14

$$\text{Standard deviation (SD)} = 7.65$$

$$\text{Standard Deviation Reduction (SDR)} = \left( \frac{5}{14} \times 10.87 \right) + \left( \frac{5}{14} \times 7.78 \right) + \left( \frac{4}{14} \times 3.4910 \right) = SD$$

$$SDR = 9.3210 - SD$$

$$SDR = 1.66215$$

Temperature	Mean	SD	Count	$w(T)$
Hot	36.25	8.954	4	4/14
Cool	39	10.511	4	4/14
Mild	42.666	7.652	6	6/14

$$\text{Standard deviation (SD)} = \left( \frac{4}{14} \times 8.954 \right) + \left( \frac{4}{14} \times 10.511 \right) + \left( \frac{6}{14} \times 7.652 \right) = 8.8413$$

$$SD = 8.8413$$

$$SDR = 0.4796$$

Humidity	Mean	SD	Count	$w(H)$
High	37.5714	9.3634	7	7/14
Normal	42	8.7341	7	7/14

$$SD = 9.0487$$

$$SDR = \left( \frac{7}{14} \times 9.3634 \right) + \left( \frac{7}{14} \times 8.7341 \right) = 9.0487$$

$$SDR = 0.2722$$

Windy	Mean	SD	Count	$w(W)$
TRUE	37.666	10.5934	6	6/14
FALSE	41.375	7.8730	8	8/14

$$SD = 9.03893$$

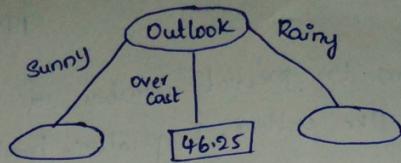
$$SDR = 0.2821$$

55

10

100%  
100%

The SDR of Outlook Attribute is high so it will be the root node of decision tree.



Here the overcast has reached decision <sup>leaf</sup> node since the termination criteria is matched (i.e.)  $CV \leq 10\%$ . and  $n=4$

Leaf node identification.

$$SD(\text{Hours}) = 10.8701426$$

Outlook : Sunny

Temperature	Mean	SD	Count
Cool	37.5	14.5	2
Mild	40.33	7.3181	3

$$\text{Standard Deviation (SD)} = 10.19089$$

$$\text{Standard Deviation Reduction (SDR)} = 0.6792$$

Humidity	Mean	SD	Count
High	37.5	7.5	2
Normal	40.33	12.4988	3

$$SD = 10.4993$$

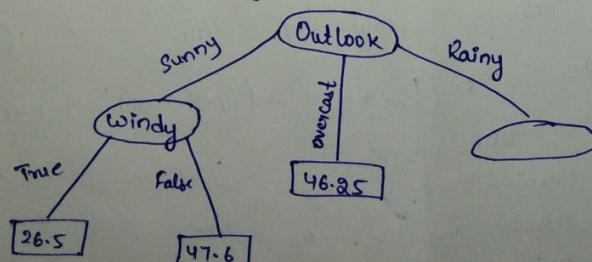
$$SDR = 0.37080$$

Windy	Mean	SD	Count
True	26.5	3.5	2
False	47.66	3.09	3

$$SD = 3.25472$$

$$SDR = 7.6154$$

→ The SDR of windy is maximum. So it will be the decision node.



→ The windy attribute has two leaf nodes which reached the termination criteria of  $n=4$  and  $CV \leq 10\%$ .

Outlook : Rainy

19K41A0555

③

Temperature	Mean	SD	Count
Hot	27.5	2.5	2
Cool	38	0	1
Mild	41.5	6.5	2

$$SD = 3.6$$

$$SDR = 4.182030$$

Humidity	Mean	SD	Count
High	30	4.0824	3
Normal	43	5	2

$$SD = 4.4494$$

$$SDR = 3.83254$$

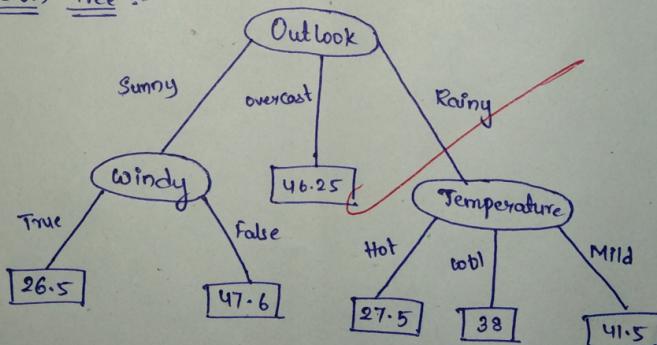
Windy	Mean	SD	Count
True	39	9	2
False	32.666	5.5577	3

$$SD = 6.93467$$

$$SDR = 0.847$$

The SDR of temperature is high. So, it will be the next decision node.

Decision Tree :-



The above Decision Tree is terminated with all leaf nodes that satisfies the termination criteria ( $n=4$ ,  $cv \leq 10\%$ .)

J. Jinaly