Date.....

Information Gain

Q What is Information Gain?

Ans Information Gain is a metric used to train Decision Trees. Specially, this metric measures the quality of the split. The Information Gain is based on the decrease in entropy after a data set is split on an attribute. Constructing a decision tree is all about finding out the attribute that returns the highest information gain.

Q What is the formula to calculate the Information Gain?

Ans **Information Gain = E(Parent) − Weighted Average * E(Child)**

Example:-

Outlook	Temperature	Humidity	Windy	Play Tennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Play_Tennis_Data Set

Date.....

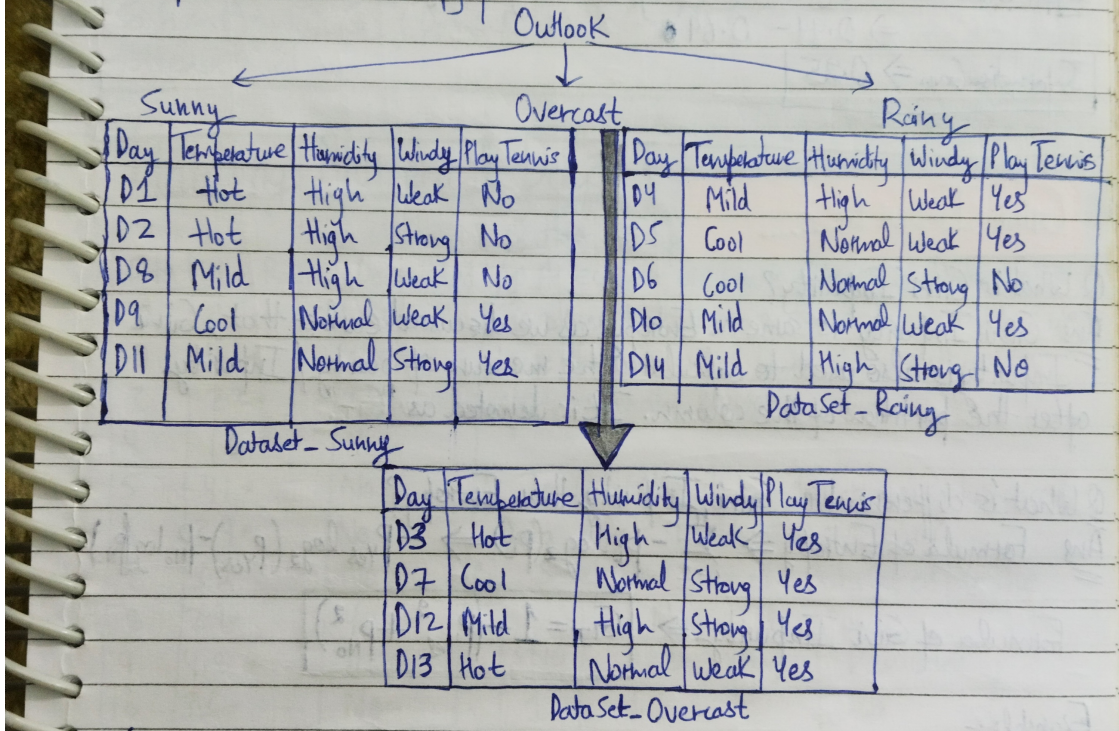
Step-1:- Calculating the E (Parent)

$\Rightarrow E(P) \Rightarrow -P_{Yes} \log_2(P_{Yes}) - P_{No} \log_2(P_{No})$

$\Rightarrow -9/14 \log_2(9/14) - 5/14 \log_2(5/14)$

$\boxed{E(P) \Rightarrow 0.94}$ - (1)

Step-2:- Calculate Entropy for Children :-



Day	Temperature	Humidity	Windy	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

Dataset_Sunny

Day	Temperature	Humidity	Windy	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
D10	Mild	Normal	Weak	Yes
D14	Mild	High	Strong	No

DataSet_Rainy

Day	Temperature	Humidity	Windy	Play Tennis
D3	Hot	High	Weak	Yes
D7	Cool	Normal	Strong	Yes
D12	Mild	High	Strong	Yes
D13	Hot	Normal	Weak	Yes

DataSet_Overcast

$E(D\_Sunny) \Rightarrow -P_{Yes} \log_2(P_{Yes}) - P_{No} \log_2(P_{No})$

$\Rightarrow -2/5 \log_2(2/5) - 3/5 \log_2(3/5)$

$\Rightarrow \boxed{0.97}$ - (2)

$E(D\_Overcast) = -P_{Yes} \log_2(P_{Yes}) - P_{No} \log_2(P_{No})$

$\Rightarrow 4/4 \log_2(4/4) - 0$

$\Rightarrow \boxed{0}$ - (3)

$E(D\_Rainy) \Rightarrow -P_{Yes} \log_2(P_{Yes}) - P_{No} \log_2(P_{No})$

$\Rightarrow -3/5 \log_2(3/5) - 2/5 \log_2(2/5)$

$\Rightarrow \boxed{0.97}$ - (4)

Spiral

Date.....

Step-3:- Calculate the Weighted Entropy of Children...

Weighted Entropy $\Rightarrow 5/14 * 0.97 + 4/14 * 0 + 5/14 * 0.97$

$\Rightarrow 0.3464 + 0.3464$

W.E (Children) $\Rightarrow 0.69$ -(5)

Taking (1), (2), (3), (4) and (5)

Information Gain = E(Parent) - Weighted Average * E(Children)

$\Rightarrow 0.94 - 0.69$

Information Gain $\Rightarrow 0.25$