Decision Tree :- Decision tree is a supervised learning technique that can be used for both classification and regression problems but mostly it is preferred for solving classification problems. it is a tree structured classifier, where internal nodes represent the features of dataset, branches represent the decision rule and each leaf node represent the outcome.

why use decision tree?

#Decision trees usually mimic human thinking ability while making a decision so it is easy to understand.

the logic behind the decision tree can be easily understood because it shows a tree like structure.

Outlook	Temperature	Routine	Wear Jacket
Sunny	cold	Indoor	No
Sunny	warm	outdoor	No
cloudy	warm	Indoor	No
sunny	warm	Indoor	No
cloudy	cold	Indoor	Yes
cloudy	cold	outdoor	Yes
sunny	cold	outdoor	Yes

m

$E(S) = \sum_{i=1} p_i \log_2 p_i$

$IG(Y;X) = E(Y) - E(Y|X)$

Gini index $= 1 - \sum_{i=1}^{n} p_i^2$

⇒ Entropy

⇒ Information gain

⇒ integer number :-

$\log_2 4$

$\Rightarrow \log_2 2^2$

$\Rightarrow 2 \log_2 2$

⇒ 2 Ans

or,

$\log_2 4$

$\Rightarrow \frac{\log 4}{\log 2}$

= 2 Ans

$\log_2 2 = 1$

Base change rule :-

$\log_b a = \frac{\log a}{\log b}$

(ii) Fraction numbers:-

$\log_2\left(\frac{1}{4}\right) \Rightarrow \frac{\log\left(\frac{1}{4}\right)}{\log 2}$ | $\log\left(\frac{a}{b}\right) = \log a - \log b$

$\Rightarrow \frac{\log 1 - \log 4}{\log 2}$

$\Rightarrow \frac{0 - \log 2^2}{\log 2}$

$\Rightarrow \frac{-2 \log 2^2}{\log 2}$

$= -2$ Ans

$\overline{}$
4

	wear	Jacket
1	Yes	3 times
2	No	4 times

Find Entropy :- $E(Y) = \sum_{i=1} p_i \log_2 p_i$

$E(Y)$ = Entropy Before Partition

$E(Y|X)$ = Entropy After Partition

Target $E(Y) >> E(Y|X)$

$\Rightarrow$ find Entropy 2

(4, 3)

$\Rightarrow (-(p_i \log_2 p_i) + (-p_i \log_2 p_i))$

$\Rightarrow \left(-\frac{4}{7} \cdot \log_2 \frac{4}{7}\right) + \left(-3/7 \log_2 3/7\right)$

$\Rightarrow$ ~~0.70 + 0.45~~

$\Rightarrow (-0.57 \log_2 0.57) + (-0.42 \log_2 0.42)$

$\Rightarrow 0.98 = E(Y)$

4

	wear	Jacket
1	Yes	3 times
2	No	4 times

Find Entropy :- $E(y) = \sum_{i=1} p_i \log_2 p_i$

$E(Y)$ = Entropy Before Partition

$E(Y|X)$ = Entropy After Partition

Target $E(Y) >> E(Y|X)$

$\Rightarrow$ find Entropy 2

$(4, 3)$

$\Rightarrow \left(-(p_i \log_2 p_i) + (-p_i \log_2 p_i)\right)$

$\Rightarrow \left(-\frac{4}{7} \cdot \log_2 \frac{4}{7}\right) + \left(-3/7 \log_2 3/7\right)$

$\Rightarrow$ ~~0.70 + 0.44~~

$\Rightarrow \left(-0.57 \log_2 0.57\right) + \left(-0.42 \log_2 0.42\right)$

$\Rightarrow 0.98 = E(Y)$

p

Outlook	
sunny	4
cloudy	3

E (outlook, sunny)

(yes) (no)

$\Rightarrow -\left(\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right)$

$\Rightarrow 0.812$

E (outlook, cloudy)

(yes) (no)

$\Rightarrow -\left(\frac{2}{3}\log_2 2/3 + \frac{1}{3}\log_2\frac{1}{3}\right)$

$\Rightarrow 0.918$

info gain $\Rightarrow E(y) - (4/7 * 0.812) - (3/7 * 0.918)$

$= 0.127$

Outlook = 0.127

Temperature	
Cold	4
warm	3

E (temperature, cold)

[illegible]

$$\Rightarrow -\left(\frac{3}{4}\log_2 3/4 + \frac{1}{4}\log_2 \frac{1}{4}\right)$$

$$\Rightarrow -(0.75\log_2 0.75 + 0.25\log_2 0.25)$$

$$= 0.811$$

E (Temperature, warm) =

$$\Rightarrow -\left(\frac{3}{3}\log_2 3/3 + \frac{0}{3}\log_2 \frac{0}{3}\right)$$

$$0.00$$

information gain $\Rightarrow E(S) - \left(\frac{4}{7} * 0.811\right)$

$$-(3/7 * 0.00)$$

E(S) = 0.98

$\Rightarrow 0.520$ Ans

Routine	
indoor	4
outdoor	3

⇒ E (Rotetine, Indoor)

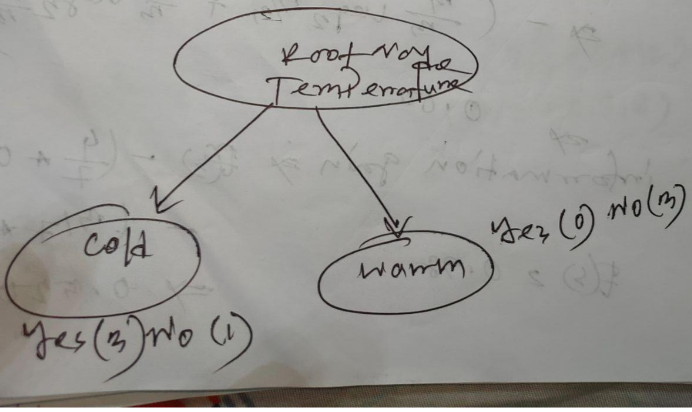
$$\Rightarrow -\left(\overset{(\text{yes})}{\frac{1}{4}\log_2 \cancel{\log}\frac{1}{4}} + \overset{(\text{no})}{\frac{3}{4}\log_2\frac{3}{4}}\right)$$

⇒ 0.812

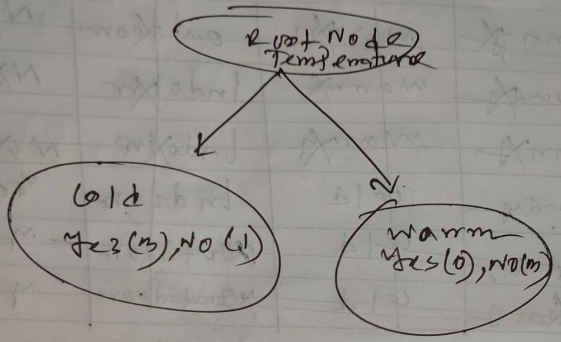
⇒ E (Routine, Outdoor)

$$\Rightarrow -\left(\overset{(\text{yes})}{\frac{2}{3}\log_2\frac{2}{3}} + \overset{(\text{no})}{\frac{1}{3}\log_2\frac{1}{3}}\right)$$

⇒ 0.127



Gini index $= 1 - \sum_{i=1}^{n} p_i^2$:



gini index $= 1 - \left(\frac{0}{3}\right)^2 + \left(\frac{3}{3}\right)^2$

~~⇒ 1 - 0 + 1~~

For warm

~~⇒ 0~~

⇒ $1 - 0 - 1$

$= 0$

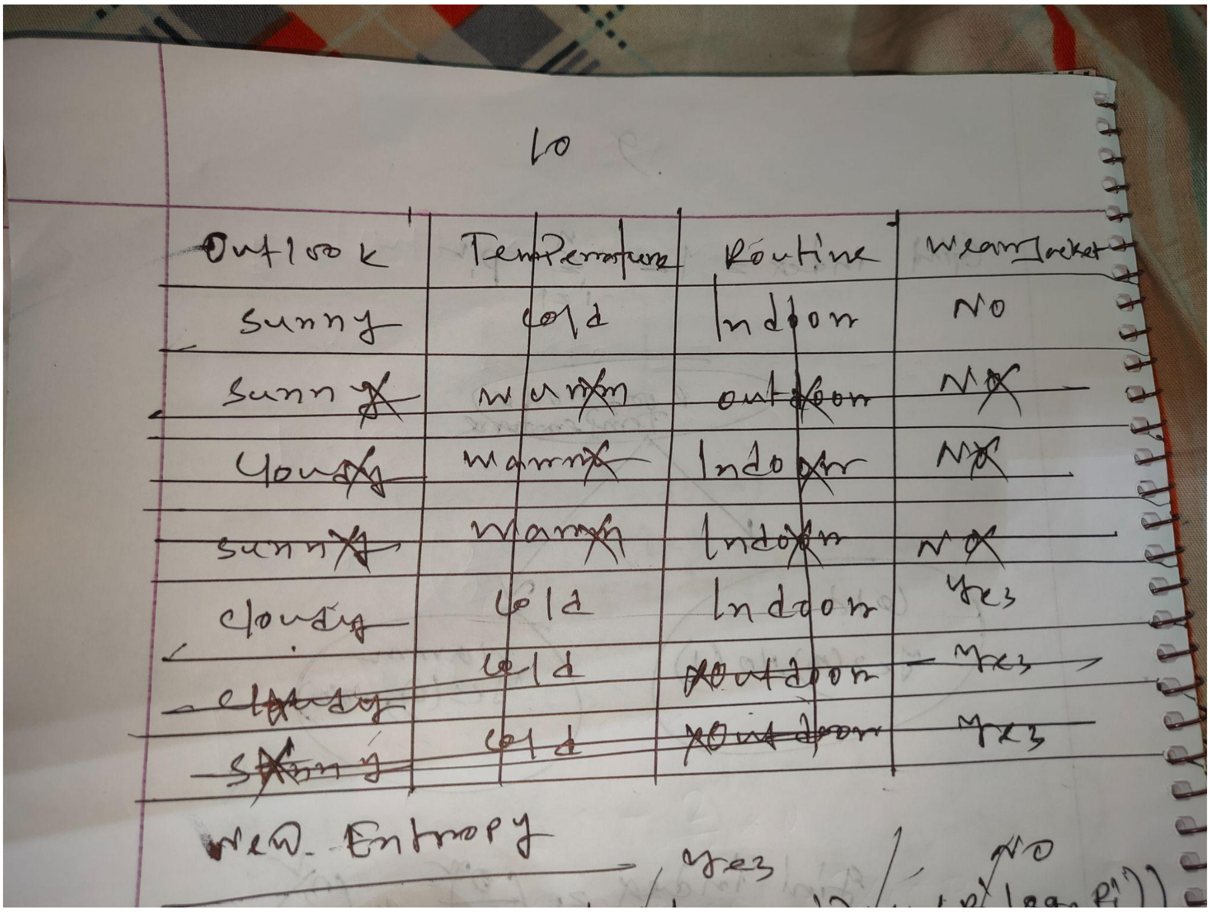
For cold

⇒ $1 - \left\{\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right\}$

$1 - \left\{(0.75)^2 + (0.25)^2\right\}$

$1 - (0.5625 + 0.0625)$

$= 0.375$



10

Outlook	Temperature	Routine	WearJacket
sunny	cold	Indoor	No
~~sunny~~	~~warm~~	~~outdoor~~	~~No~~
~~cloudy~~	~~warm~~	~~Indoor~~	~~No~~
~~sunny~~	~~warm~~	~~Indoor~~	~~No~~
cloudy	cold	Indoor	Yes
~~cloudy~~	~~cold~~	~~outdoor~~	~~Yes~~
~~sunny~~	~~cold~~	~~outdoor~~	~~Yes~~

New_ Entropy

Yes / No

New Entropy:-

yes no

$$S2 = (-p_i \log_2 p_i) + (-p_i \log_2 p_i)$$

$$= \left(-\frac{3}{4} \log_2 \frac{3}{4}\right) + \left(-\frac{1}{4} . \log_2 \frac{1}{4}\right)$$

$$= \left(-0.75 \log_2 \frac{3}{4}\right) + \left(-0.25 \log_2 \frac{1}{4}\right)$$

$$= 0.311 + 0.5$$

$$= 0.811$$

⇒ E (Routine, Indoor)

$$\Rightarrow -\left(\overset{(yes)}{\frac{1}{2}} \log_2 \frac{1}{2} + \overset{(No)}{\frac{1}{2}} \log_2 \frac{1}{2}\right)$$

$$= 1$$

E (Routine, Outdoor)

$$\Rightarrow -\left(\overset{(yes)}{\frac{2}{2}} \log_2 \frac{2}{2} + \overset{no}{\frac{0}{2}} \log_2 \frac{0}{2}\right)$$

$$\Rightarrow 0$$

Information gain:-

$$E(S2) - \frac{2}{4} * 1 - \frac{2}{4} * 0$$

$$\Rightarrow 0.811 - \frac{2}{4} * 1 - \frac{2}{4} * 0$$

$$= 0.312$$

E (outlook, sunny)

$\Rightarrow -\left(\overset{yes}{\frac{1}{2}\log_2\frac{1}{2}} + \overset{no}{\frac{1}{2}\log_2\frac{1}{2}}\right)$

$\Rightarrow 1$

E (outlook, cloudy)

$\Rightarrow -\left(\frac{2}{2}\log_2\frac{2}{2} + \frac{0}{2}\log_2\frac{0}{2}\right)$

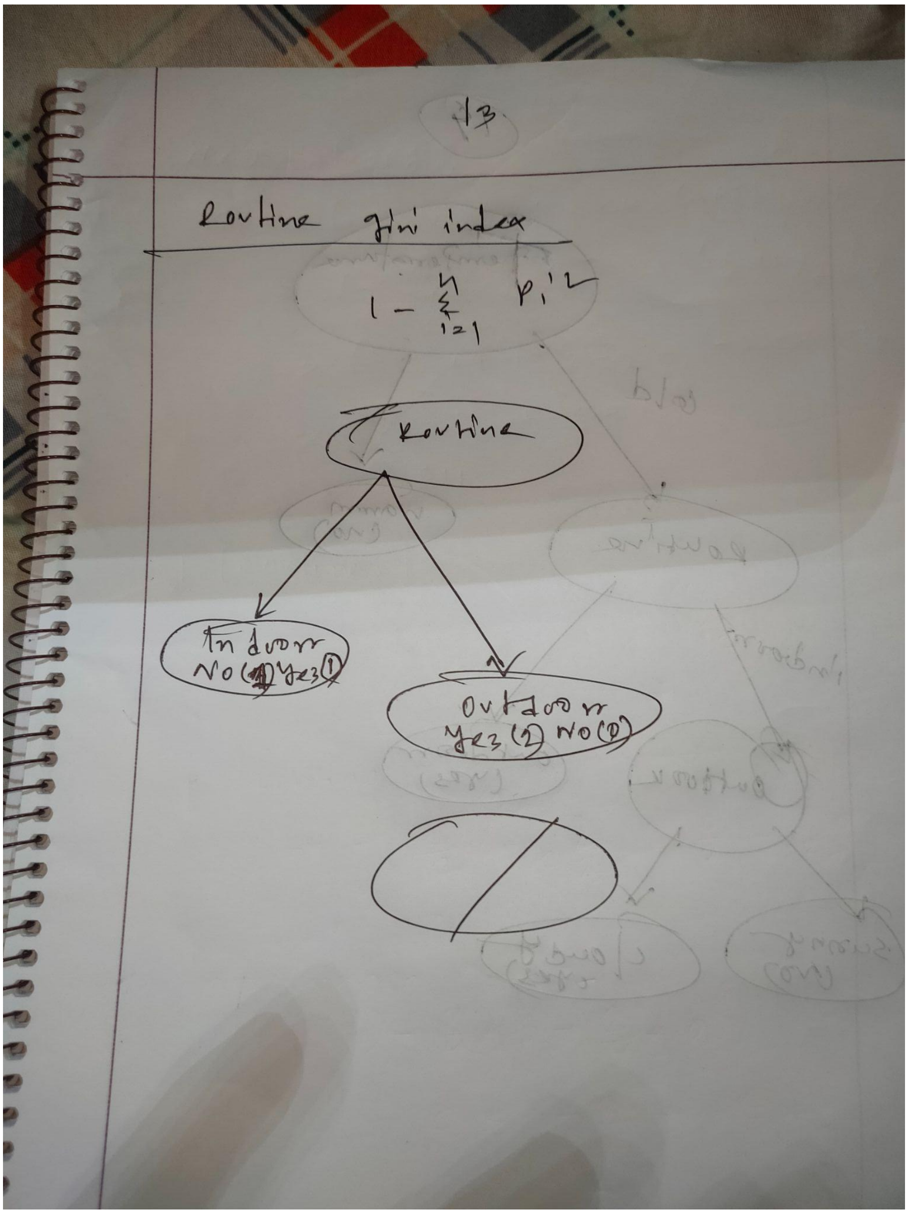
$\Rightarrow 0$

Information gain

$E(32) - \frac{2}{4} * 1 - \frac{2}{4} * 0$

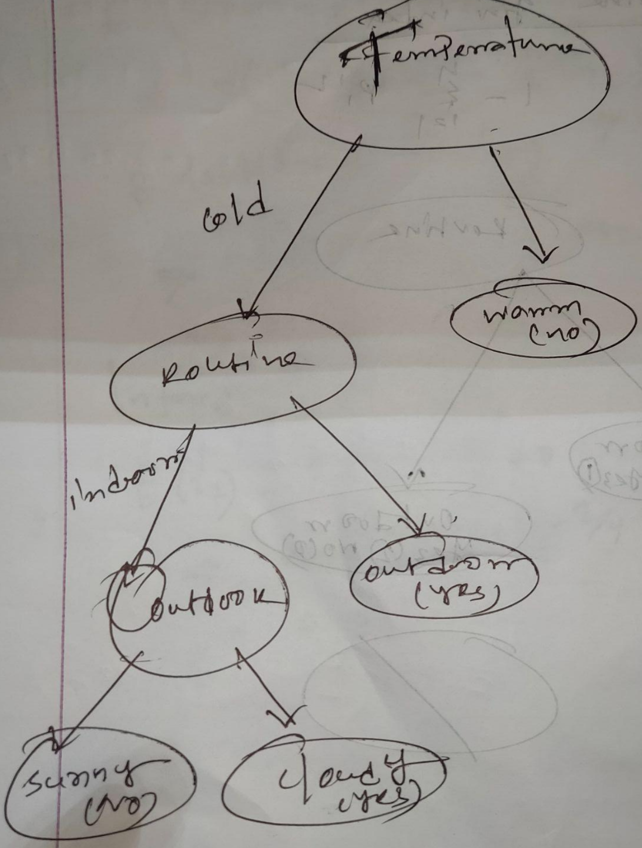
$\Rightarrow 0.811 - 2/4 * 1 - 2/4 * 0$

$\Rightarrow 0.312$



Routine gini index

$$1 - \sum_{i=1}^{n} p_i^2$$


Temperature
cold
warm
(no)
Routine
indoor
outdoor
(yes)
outlook
sunny
(no)
cloudy
(yes)