①

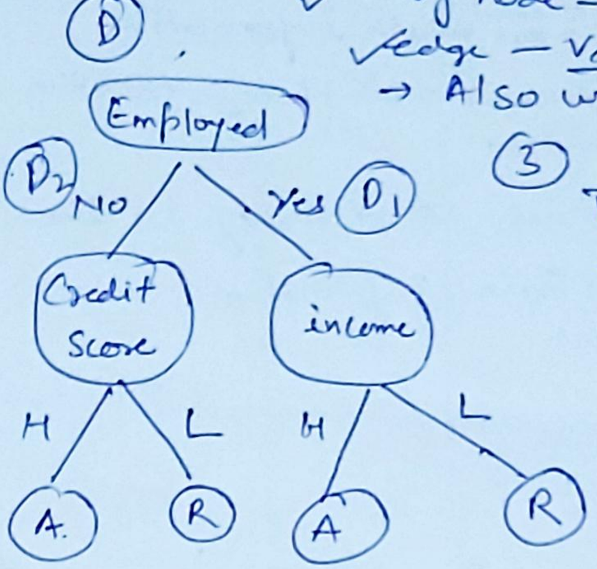
Decision tree :- ✓A classifier (Tree Structured), ① Decision Node (Test) /feature /Attribute

②

✓Leaf node – (Classification/~~Value~~)

✓edge – Value

→ Also used in Regression



③ Training Data — Algo → m/c

Generate

↑ tuple (Test)

→ Loan or not prediction

→ Test is performed on the feature/Attribute

Train DS

④ D

Employed	Credit score	income		Status – Target / Leaf Node
Y	H	H		A
Y	H	H	D1	A
Y	H	H		A
H	L	L	D2	R
H	L	L		R

further do the splitting / Leaf node having class or Value.

Question:1:- for the following Medical diagnostic data create DT.

~~$I.G = \frac{P}{P+N}$~~

Information Gain:- Measure of how much information the Answer to a specific Question provides.

Entropy: measure of how much uncertainty in the dataset / information Gain / Information.

Info gain ↑ = entropy ↓

Information Gain:-

$$I(P,n) = -\frac{P}{S}\log_2\frac{P}{S} - \frac{n}{S}\log_2\frac{n}{S} \quad \text{— (I)}$$

S → total sample space

$S = P + N$

$$E(A) = \sum_{i=1}^{v} \frac{P_i + n_i}{P+n} \left(I(P_i, n_i)\right) \quad \text{— (II)}$$

$$Gain(A) = I(P,n) - E(A) \quad \text{— (III)}$$

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2}$$

Question :1: - for the following Medical diagnosis data, Create DT -

	Sore Throat	Fever	Swollen Glands	Congestion	Headache	Diagnosis
1	Yes	Yes	Yes	Yes	Yes	strep throat
2	NO	NO	NO	Yes	Yes	Allergy
3	Yes	Yes	NO	Yes	NO	Cold
4	Yes	NO	Yes	NO	NO	S.T
5	NO	Yes	NO	Yes	NO	Cold
6	NO	NO	NO	Yes	NO	Allergy
7	NO	NO	Yes	NO	NO	S.T
8	Yes	NO	NO	Yes	Yes	Allergy
9	NO	Yes	NO	Yes	Yes	Cold
10	Yes	~~NO~~ Yes	NO	Yes	Yes.	Cold.

Sample space :- $S.T + All + Cold = 10$

$$\begin{array}{ccc} S.T & All & Cold \\ \downarrow & \downarrow & \downarrow \\ 3 & 3 & 4 \end{array}$$

~~I(P,n)~~

$$I(ST, All, Cold) = -\left[\left(\frac{3}{10}\right)\log_2\left(\frac{3}{10}\right) + \frac{3}{10}\log_2\left(\frac{3}{10}\right) + \left(\frac{4}{10}\right)\log_2\left(\frac{4}{10}\right)\right]$$

$$= \underline{1.562}$$

finding Splitting Attribute :. [Select Attribute with Highest gain]

Sore Throat : —

	ST	A	C	
Yes	2	1	2	(Infogain) × P
NO	1	2	2	(Info gain) × P

+ E(A)

E(Sore Throat)

$$I(\text{Yes}) = -\left[\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \left(\frac{1}{5}\right)\log_2\left(\frac{1}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right]$$

$I(\text{Yes}) = 1.52$

$I(\text{NO}) = 1.52$

$$E(\text{Sore throat}) = \frac{5}{10} \times 1.52 + \frac{5}{10} \times 1.52 = 1.52$$

$\text{Gain}(\text{Sore throat}) = 1.562 - 1.52 = 0.05$

Sore Throat — 0.05

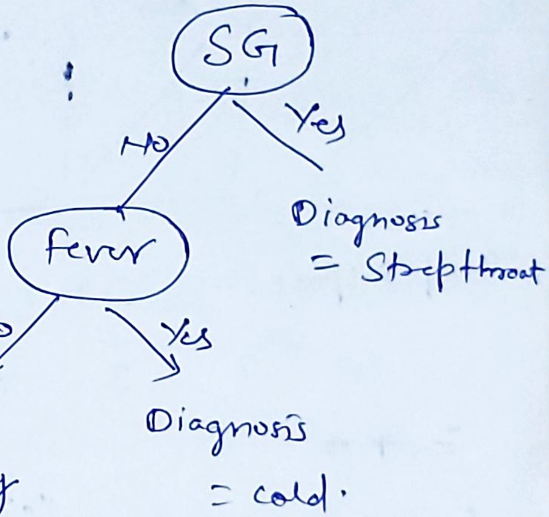
fever → $\boxed{0.72}$

Swollen glands → 0.88].

Congestion — 0.45

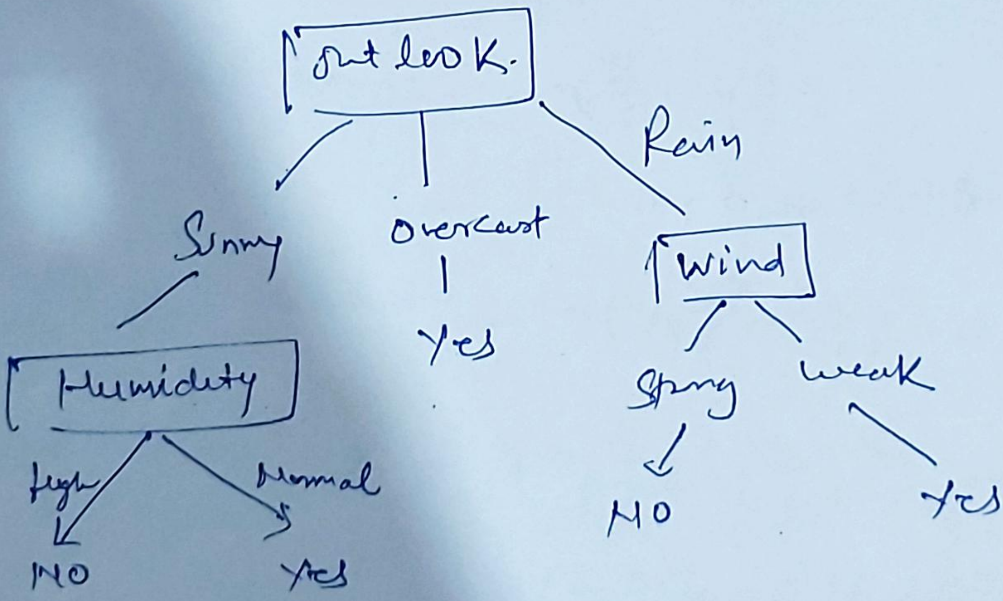
Headache → 0.05

Decision Tree



Create a Decision tree for given following Data:-

Day	Out look	Tem	Humidity	wind	Play.
1	Sonny	Hot	High	weak	No
2	Sunny	"	"	Strong	No
3	Overcast	"	"	weak	Yes
4	Rain	mild	"	"	"
5	Rain.	cold	Normal	"	"
6	Rain	"	"	Strong	No
7	overcast	"	"	"	Yes
8	Sonny	mild	High	weak	No
9	"	cold	Normal	"	Yes
10	Rain	mild	"	"	Yes
11	Sonny	"	"	Strong	Yes
12	overcast	"	High	"	"
13	"	Hot.	Normal	weak	"
14	Rain	mild	High	Strong	No



CART → Gini(S) = 1 −

entropy
↓
Gini

$$Gini(E) = 1 - \sum_{j=1}^{c} p_j^{2}$$

$$= 1 -$$

← Entropy | IG

ID3 — Iterative Dichotomiser 3

CART — Classification & Regression tree.

↓

Gini Index

CART (Classification & Regression trees) - Decision tree

Gini Index:- Probability of each class:-

Sum of squared probability of each class. we can formulate it as below:-

$$Gini = 1 - \Sigma (P_i)^2 \text{ for } i = 1 \text{ to number of classes.}$$

Outlook: outlook is nominal feature

Sunny overcast rain

Outlook	Yes	No	no. of instances
Sunny	2	3	5
overcast	4	0	4
Rain	3	2	5.

$$Gini(outlook = Sunny) = 1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2$$

$$= 1 - 0.16 - 0.36 = 0.48$$

$$Gini(outlook = overcast) = 1 - (4/4)^2 - \left(\frac{0}{4}\right)^2 = 0$$

$$Gini(outlook = Rain) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2$$

$$= 1 - 0.36 - 0.16 = 0.48$$

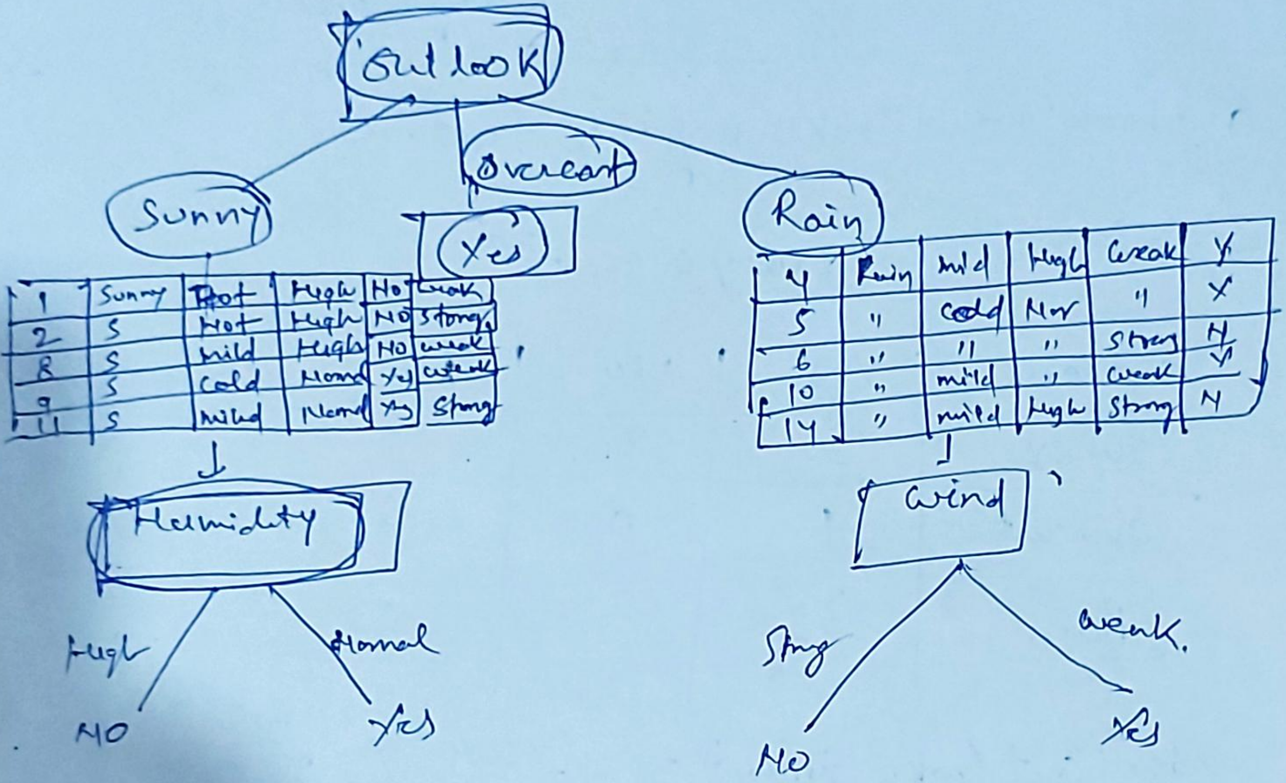
$$Gini(outlook) = \frac{5}{14} \times 0.48 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.48$$

Gini(outlook) = 0.342 ✓ lowest cost

Gini(temp) = 0.439

Gini(Humidity) = 0.367

Gini(wind) = 0.428



1	Sunny	Hot	High	NO	weak
2	S	Hot	High	NO	Strong
8	S	mild	High	NO	weak
9	S	cold	Normal	Yes	weak
11	S	mild	Normal	Yes	Strong

4	Rain	mild	High	weak	Y
5	"	cold	Nor	"	Y
6	"	"	"	Strong	N
10	"	mild	"	weak	Y
14	"	mild	High	Strong	N