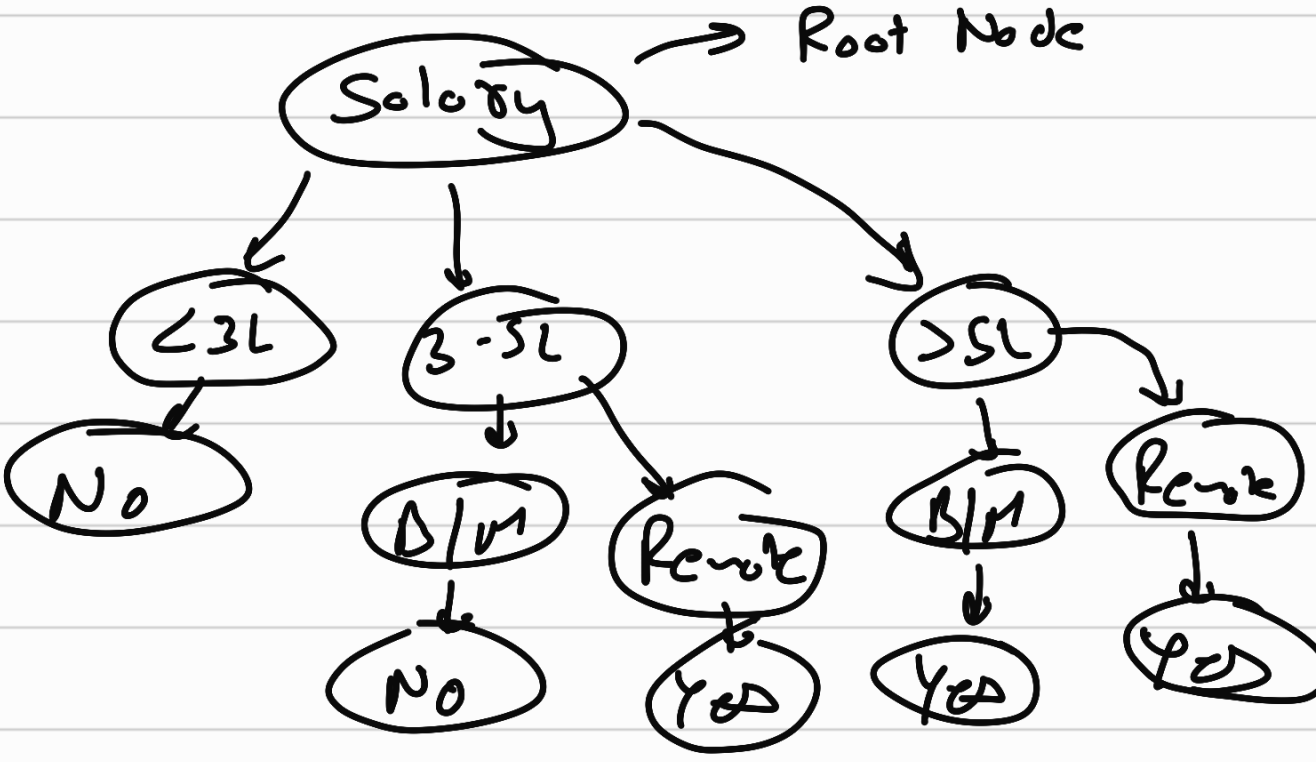
Feature Scaling:-

① Std^n → Standard Scaler → $x' = \frac{x - \mu}{\sigma}$

② Norm → Min Max Scaler → $x' = \frac{x - x_{min}}{x_{max} - x_{min}}$

Decision Tree:-

You Will take New Job?



You have to find which feature is important

Criterion ⇒ Hyper Meter

↳ gini or entropy

↓

based on this D.T decided Root Node

Eg:-

Decision Tree Algorithm

Age	Competition	Type	Profit/Loss
Old	Yes	Software	Down
Old	No	S	Down
Old	No	Hardware	Down
Mid	Yes	S	Down
Mid	Yes	H	Down
Mid	No	H	Up
Mid	No	S	Up
New	Yes	S	Up
New	No	H	Up
New	No	S	Up

⇒5

⇒5

Problem Statement: To find for a given new transaction, whether there will be profit or loss

Y Variable: Profit/Loss column

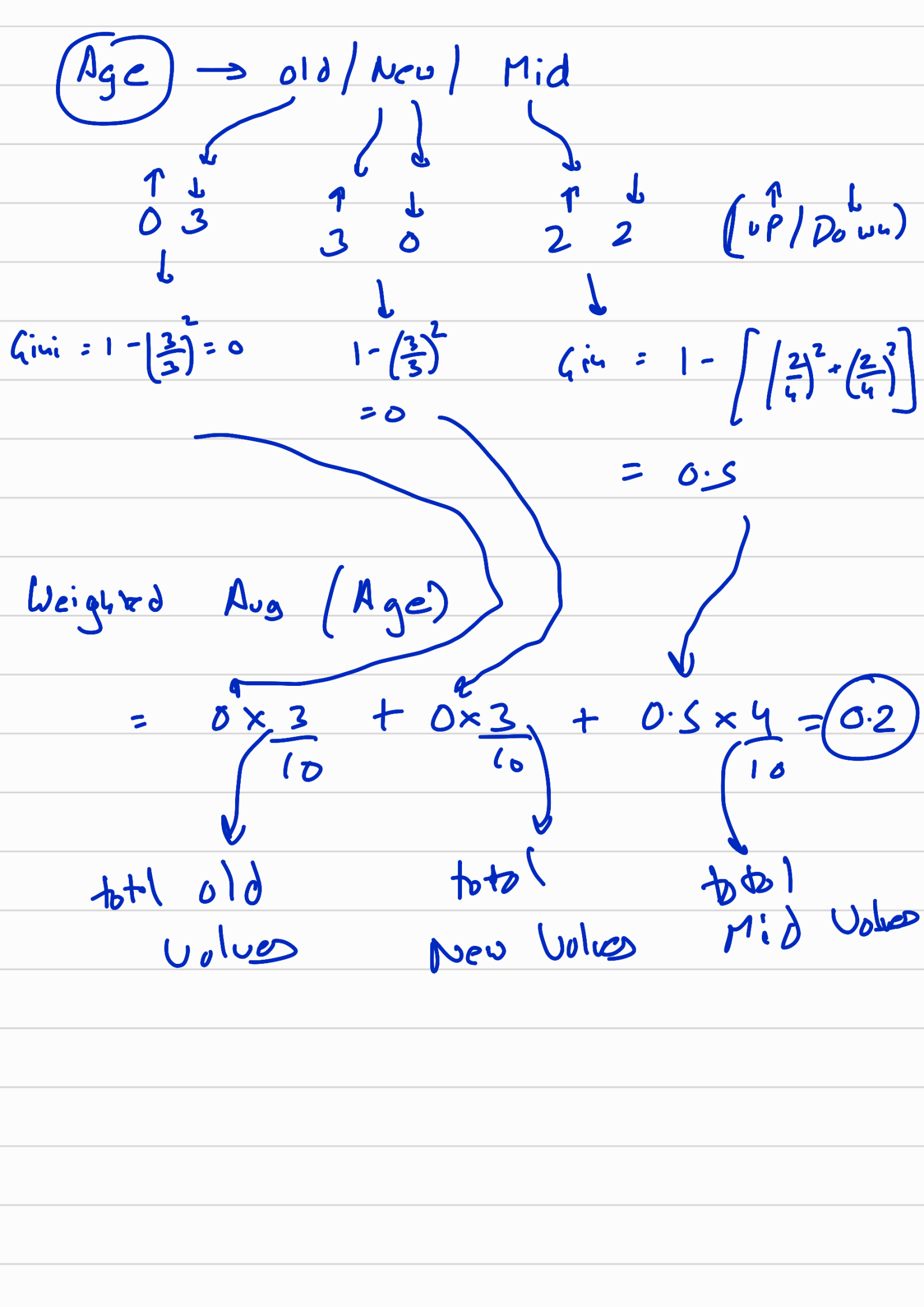
X Variables: Age, Competition, Type

We have to identify which x_i ⇒ root

Gini ↓ ⇒ More Info

$$\text{Gini}(P/L) = 1 - \left[\left(\frac{5}{10}\right)^2 + \left(\frac{5}{10}\right)^2 \right]$$

$$= 1 - [0.5] = 0.5$$


Age → old / New / Mid
↑ 0 ↓ 3
↑ 3 ↓ 0
↑ 2 ↓ 2
(↑ UP / ↓ Down)
Gini : 1 - (3/3)² = 0
1 - (3/3)²
=0
Gini = 1 - [(2/4)² + (2/4)²]
= 0.5
Weighted Avg (Age)
= 0 × 3/10 + 0 × 3/10 + 0.5 × 4/10 = 0.2
total old values
total New values
total Mid values

Competition ⇒ Yes / No

↑ 1 ↓ 3 ↑ 4 ↓ 2

$$Gini(S) = 1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2\right]$$

$$= 0.375$$

$$Gini(S) = 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right]$$

$$= 0.44$$

$$WA(C) \Rightarrow 4 \times \frac{0.375}{10} + 6 \times \frac{0.44}{10}$$

$$= 0.154 + 0.264$$

$$= 0.414$$

Type ⇒ 6 / 4

$$Gini(S) = 1 - \left[\left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right]$$

$$= 0.5$$

$$Gini(S) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right]$$

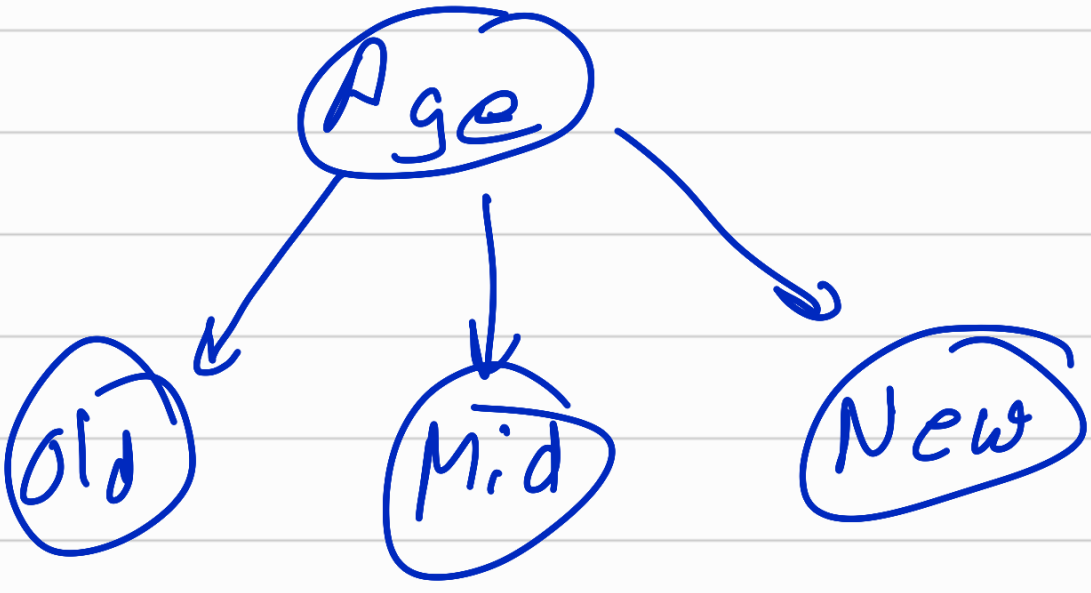
$$= 0.5$$

$$GA(Type) = \frac{6}{10} \times 0.5 + \frac{4}{10} \times 0.5$$

$$= 0.3 + 0.2$$

$$= 0.5$$

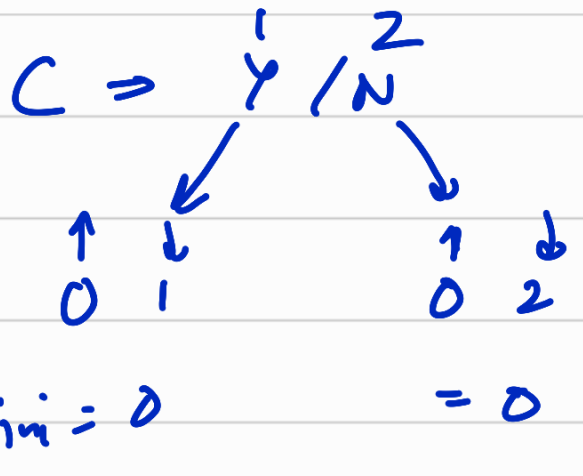
Age → 0.2 → root
C → 0.414
T → 0.5

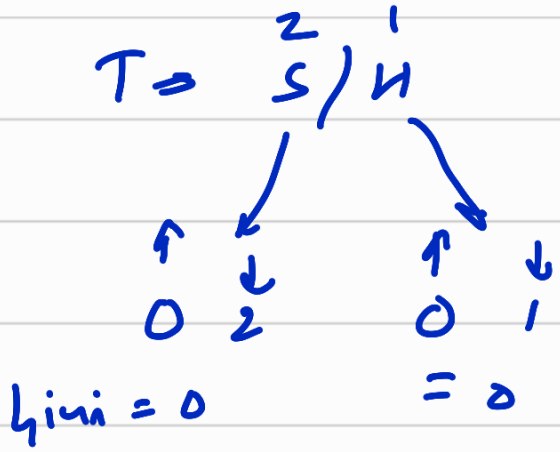


Further Calculate Again
w.r.t to old, mid & New
for Competition & Type

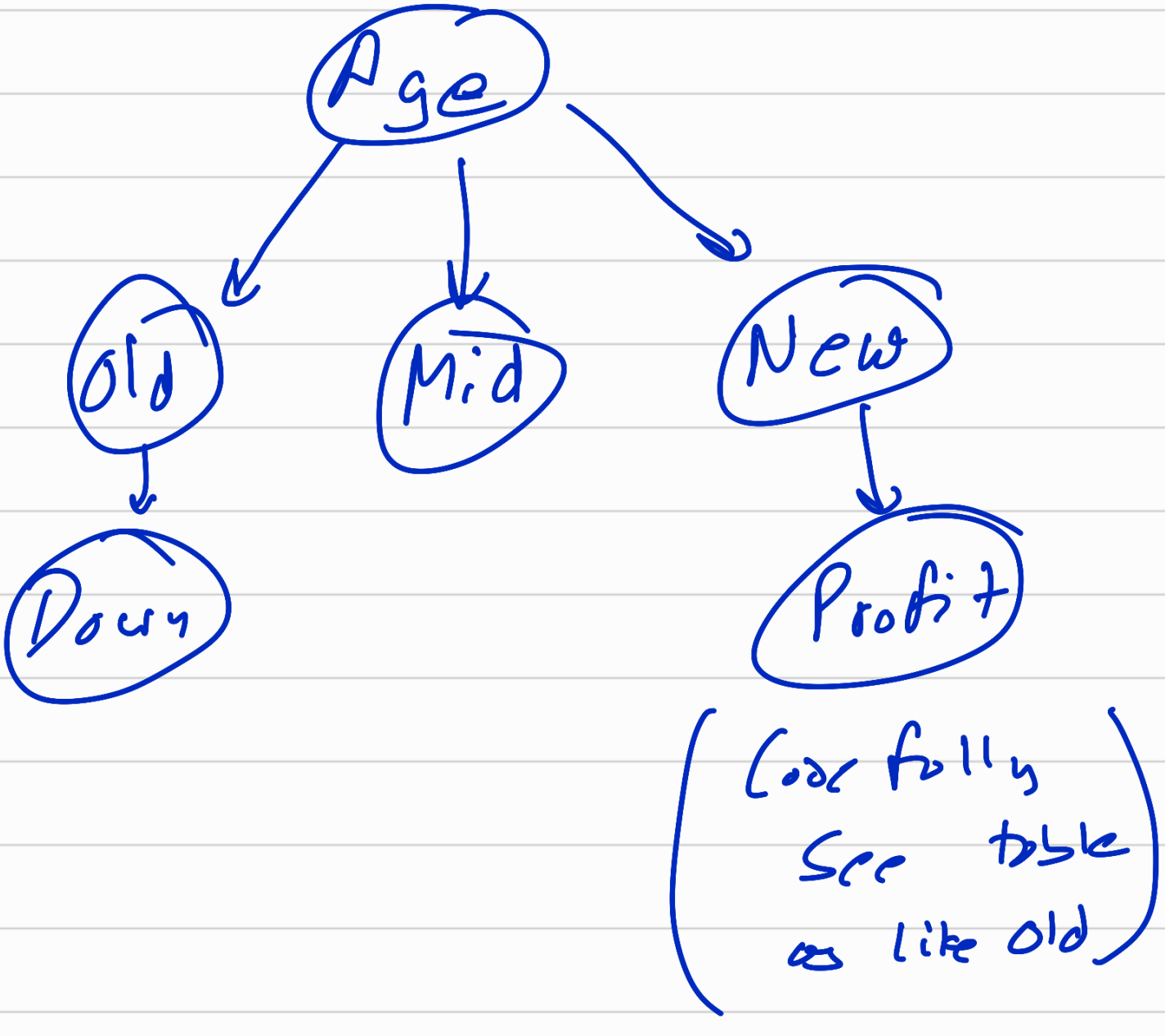
old

Age	Competition	Type	Profit/Loss
old	Yes	Software	Down
old	No	S	Down
old	No	Hardware	Down





No One is Imp so for age old you will have Loss.



Mid :-

Mid	Yes	S	Down
Mid	Yes	H	Down
Mid	No	H	Up
Mid	No	S	Up

Ignore

C → Y / N

Y: 2 (2 → 0, 0 → 2)

N: 2 (0 → 2, 2 → 0)

Gini = 0 , 0

WA(C) = 0

T → S / H

S: 2 (1 → 1, 1 → 1)

H: 2 (1 → 1, 1 → 1)

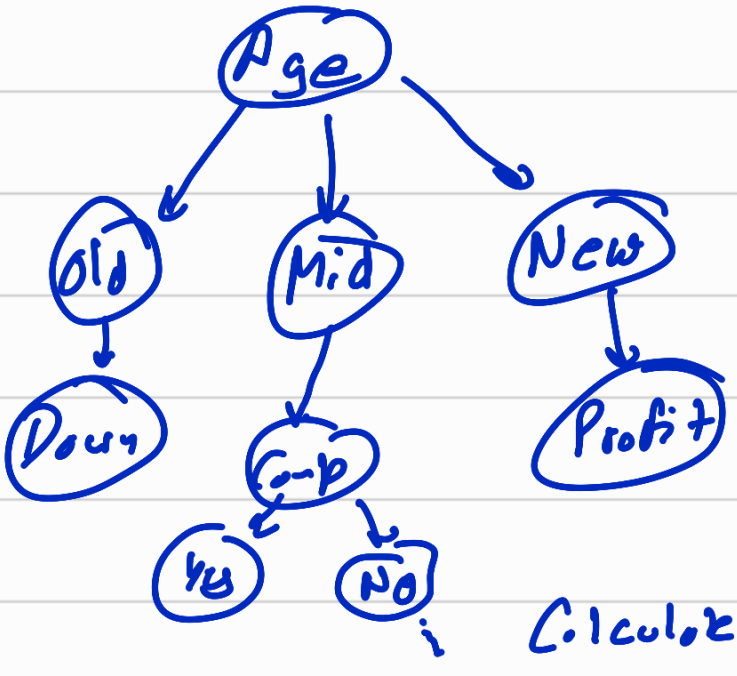
Gini: 0.5 , 0.5

$WA(T) = \frac{2}{4} \times 0.5 + \frac{2}{4} \times 0.5$

$= 0.5$

C = 0 → Lowest

T → 0.5



Calculate further

Entropy :- ET :- Info T

Decision Tree Algorithm

→ C.A

Age	Competition	Type	Profit/Loss
Old	Yes	Software	Down
Old	No	S	Down
Old	No	Hardware	Down
Mid	Yes	S	Down
Mid	Yes	H	Down
Mid	No	H	Up
Mid	No	S	Up
New	Yes	S	Up
New	No	H	Up
New	No	S	Up

P = Down (count)

N = UP (count)

Problem Statement: To find for a given new transaction, whether there will be profit or loss

Y Variable: Profit/Loss column

X Variables: Age, Competition, Type

Entropy of CA = $\frac{-P}{P+N}\log\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log\left(\frac{N}{P+N}\right)$

Class Attribute

$= \frac{-5}{10}\log\left(\frac{1}{2}\right) - \frac{5}{10}\log\left(\frac{1}{2}\right)$

$= \frac{-1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right)$

$= -1\log\left(\frac{1}{2}\right) = 1$

Age → O / M / N

	↑	↓	I.G
old	0	3	→ $\frac{-3}{3}\log\left(\frac{0}{3}\right)$
Mid	2	2	→ 1 − $\frac{-2}{3}\left(\log\frac{2}{2}\right)$
New	3	0	→ 0

= 0

$$\varepsilon_{Ag} = \frac{\Sigma P_i + N_i}{P+N} \times I \cdot G$$

$$= \frac{0+3}{10} \times 0 + \frac{2+2}{10} \times 1 + \frac{3+0}{10} \times 0$$

$$= 0.4$$

Co-petition :-

	↑	↓	I·G
Yo	1	3	0.8
No	4	2	0.9

$$\varepsilon_C = \frac{\Sigma P_i + N_i}{P+N} \times I \cdot G$$

$$= \frac{1+3}{10} \times 0.8 + \frac{4 \times 2}{10} \times 0.9$$

$$= 0.32 + 0.54 = 0.86$$

ε_{CA}
ε_{Age}
$\varepsilon_C = 0.86$

$\text{Join}(I \cdot G) = \varepsilon_{C.A} - \text{If}$

$\text{Join} = 1 - 0.4 = 0.6$

$\text{Join} = 1 - 0.86 = 0.14$

Random forest :-

↓

Combination of tree

(Combine multiple decision tree)

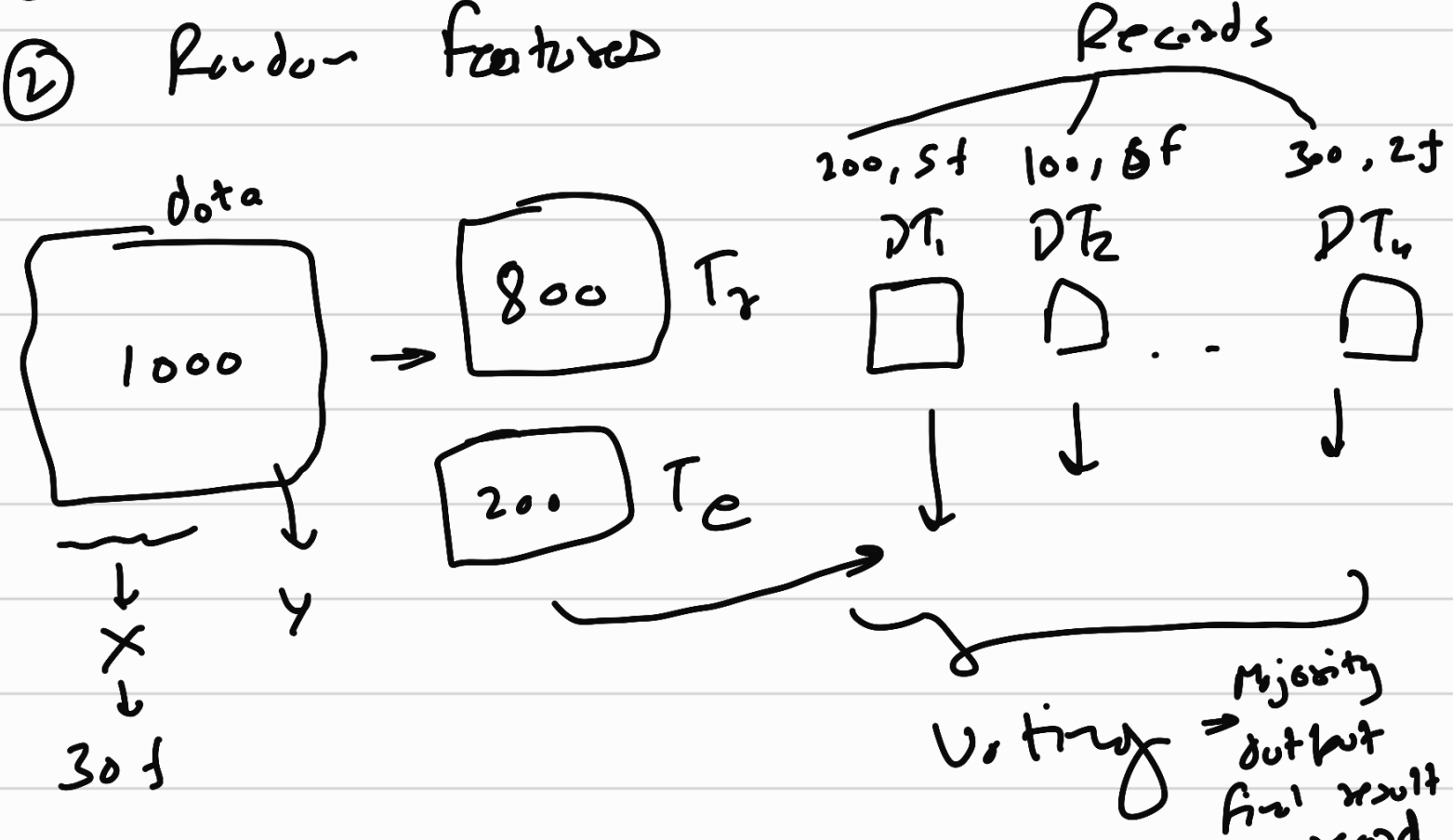
Ensemble

① Bagging → RF

② Boosting

① Random Data

② Random features



Hyper Parameter :- n_estimator, Criterion

↓

No. of tree

↓

all tree will follow some

In Most Case RF give better result than D.T but these can be exceptions.

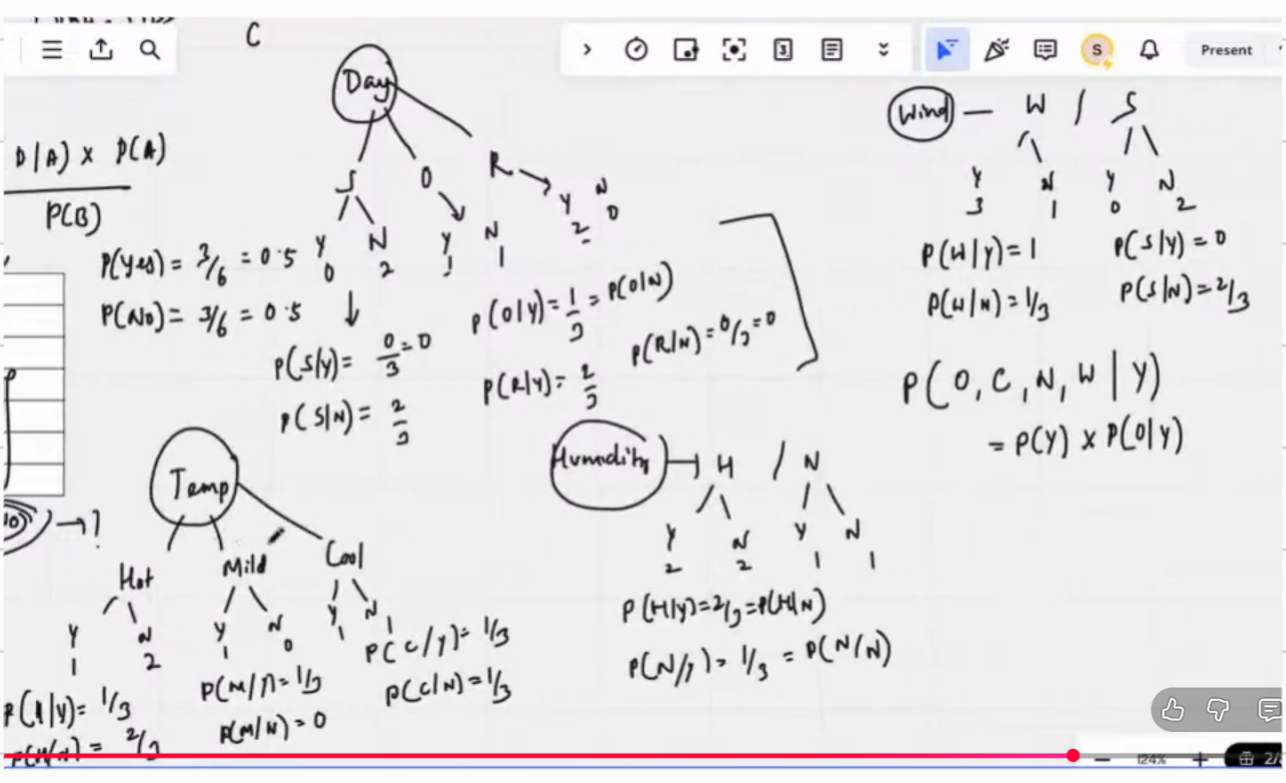
Naive Bayes :- $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

To start with, let us consider a dataset.

Day	Temp	Humidity	Wind	Play
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Overcast	Cool	Normal	Strong	No

Q - Overcast Cool Normal Weak Yes/No



$$P(O, C, N, W \mid Y)$$

$$= P(Y) \times P(O|Y) \times P(C|Y) \times P(N|Y) \times P(W|Y)$$

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{54} = 0.018$$

$$P(O, C, N, W \mid N)$$

$$= P(N) \times P(O|N) \times P(C|N) \times P(N|N) \times P(W|N)$$

$$= \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{162} = 0.006$$

Person will Play acc to N.B.