Decision trees

① Decision Trees / Classification Trees

Categorical   Continuous.
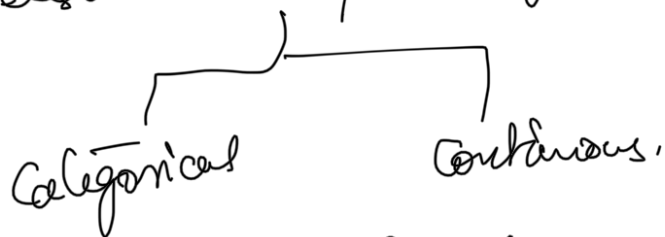
② Hands on - Lasso & Ridge.

Wed

* 1:1 Sessions
  @ 7:45 pm IST
  Wednesday

Teams :- ↳②

Sat → Next
① Decision Trees ↳ Regression Trees.
② Hands on Decision Trees.

Sun
③ Ensemble Learning ① Random Forest

Bagging Boosting

Decision Trees ⇒ Classification Trees

Love Popcorn	Love Soda	Age	Likes Movie
Yes	Yes	7	No ⟶ 0.429
Yes	No	12	No ⟶ 0.
No	Yes	18,	Yes
No	Yes	35	Yes
Yes	Yes	3t	Yes
Yes	No	30	No
No	No	70	No

Root → Loves movies / Does not love

Soda
True / False → Leaf
Criteria
Age < 2? → Not loving the movie (Leaf node)

False

True → Not loving

fall

Yes love the movie

0.405

$0.$: Age ≥ 23.5

No

Yes

Yes No

GINI Impurity
0.429

Yes No

Yes No

Root Node
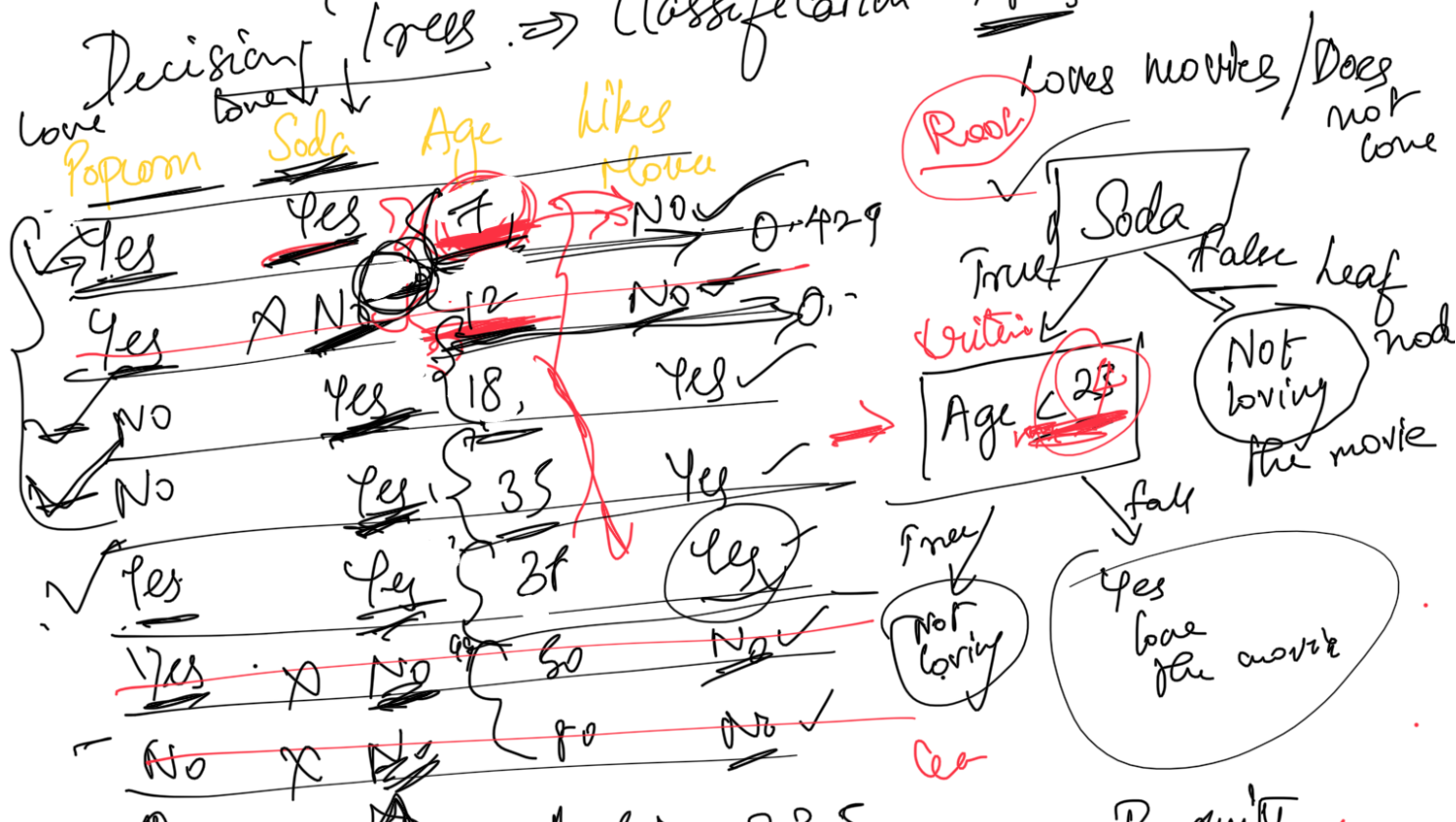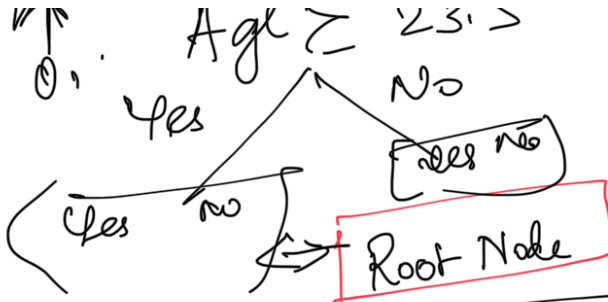
Person > 18

Person > 18

TREES

Voter's Id / Note

Branches

True

If applied

Yes No

Yes

False

No
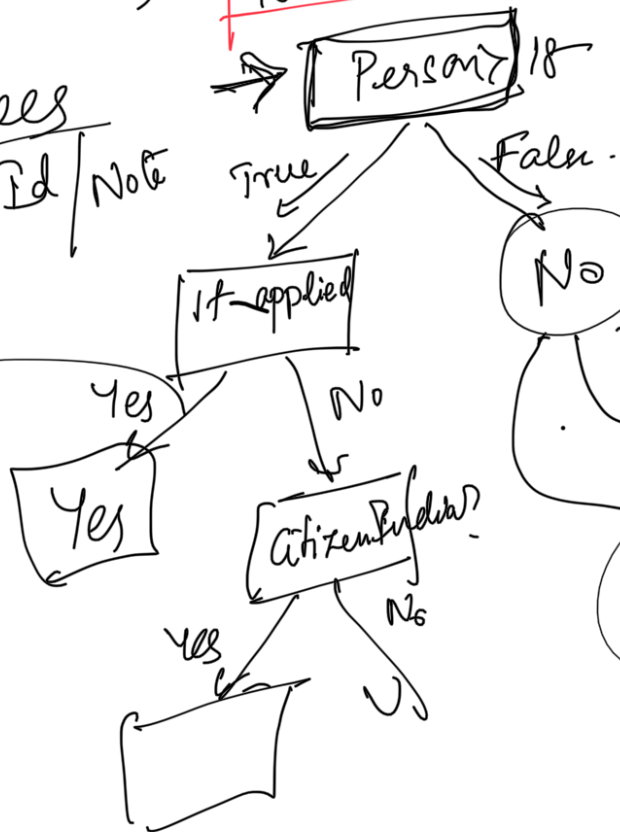
Decision
Last Node

Leaf Node

Leaf Nodes
Leaves (ov)

Citizenship?

Yes

No

Age | Loves Soda / Loves Popcorn

? ? ?

To arrive at this we create Simple tree

To
Level 0

Loves Popcorn

True   False

Loves Soda

Stumps
little
Trees

Loves Soda

True   False

Loves Magazines

Yes   No
1   3

Yes No
2   1

Yes No
3

Yes No
3   1

Yes No
3

0.425

Less Impurity
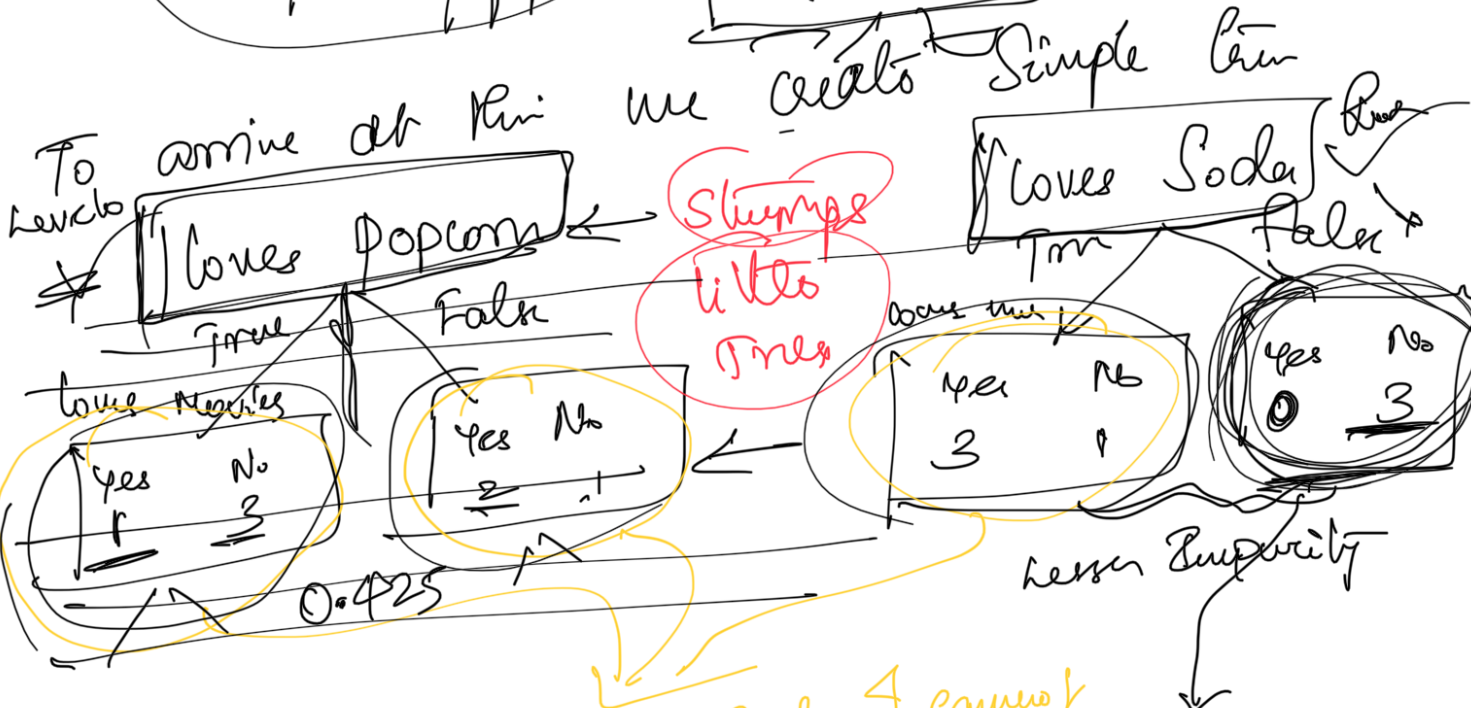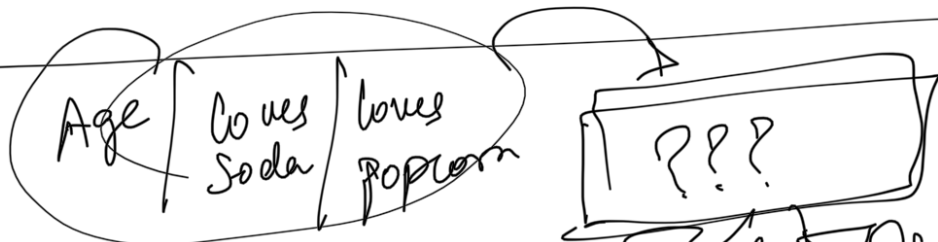
⚠️ (Impure) | Impurity is present here.

→ Impurity as a part select → free

quantify Impurity??

→① GINI Index → Impurity/Most widely use
② Entropy & Performation Gain

↳The comment present as a factor used this feel as Boam/source

Lower popu.

True   false
Yes No   Yes   No
1   3   2   1

Loves movies   Loves movies
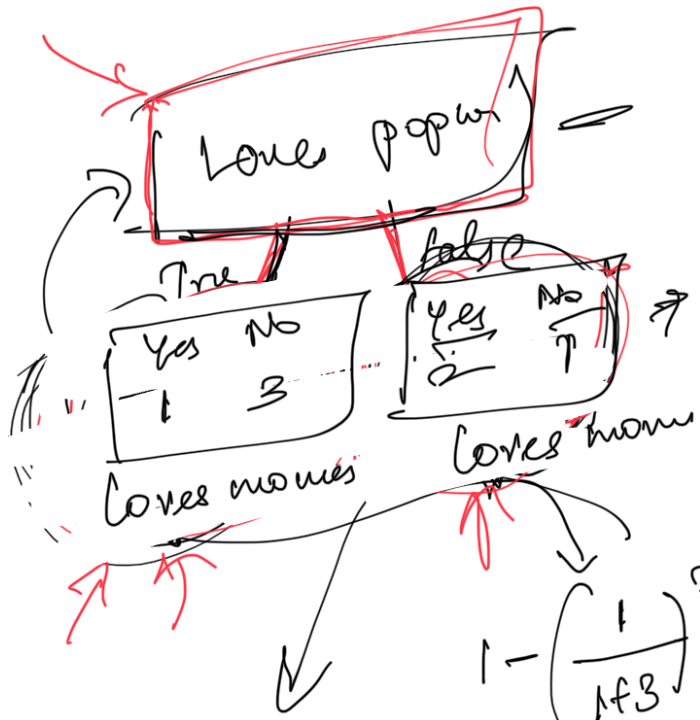
for
$$\text{Gini of the Leaf} = 1 - (\text{the probability of "Yes})^2 - (\text{the probability of "No"})^2$$

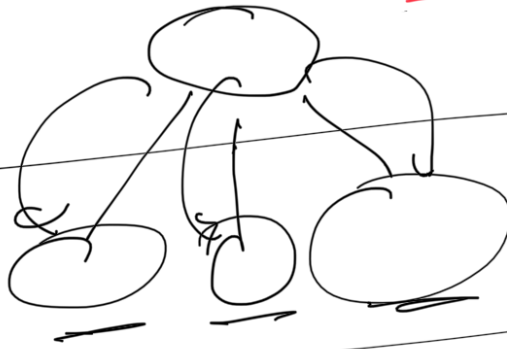$$1 - \left(\frac{1}{1+3}\right)^2 - \left(\frac{3}{1+3}\right)^2 \Rightarrow 0.375$$

$$1 - \left(\frac{2}{2+1}\right)^2 - \left(\frac{1}{2+1}\right)^2 = 0.44$$

$$\left(\frac{4}{7}\right)0.375 + \left(\frac{3}{7}\right)0.44 \Rightarrow = 0.405$$

Level 0

Level 1

& one parent
leaf

GINI Index

0.429	←	9.5
(0.393)	←	15
0.47 6	←	26.5
0.47	←	36.5
0.343	←	44
0.429	←	66.5

Age Y

	Age	Y
9.5	{ 7	No
	12	No
15	18	Yes
26.5	35	Yes
36.5	38	Yes
44	50	No
66.5	83	No

Age < 9.5

True / \ False

Yes | No   Yes | No
 0 | 4   3 | 3

$$\text{Impurity} = 1 - (P(Yes)^2) - (P(No)^2)$$

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

$$1 - (P(Yes)^2) - (P(No)^2)$$

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

$$\Rightarrow 1 - (0.25) - 0.25$$

$$= 0.5$$

Weighted Avg

$$\left(\frac{1}{1+6}\right) 0 + \frac{6}{(1+6)} \, 0.5 = 0.429$$

my Root → Least Gini Index we'll become

GINI → Measure of Impurity

GINI for loves Soda = (0.214)

min (loves popcorn, (Age mean), loves Soda)

GINI
< 15
< 9.5
< 38.5

Root Node   Loves Soda ✓

TRUE   FALSE   FINAL

Yes   No
3   Age < 12.5

Yes   No
Objective   3.   Leaf → No
caledgr
Clur

Age ?   Pop?

GINI for Soda

Loves Pop	Age	Loves Mom (Y)
Yes	F 7	No
No	P.↑ 18	Yes
No	26 < 35	Yes
Yes	36 < 88	Yes

What lowest GINI

min (Pop, Age)

min (4)

Break fell 10:20 AM 1st