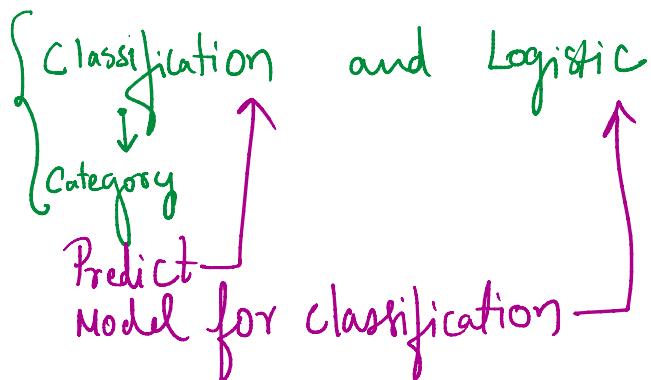
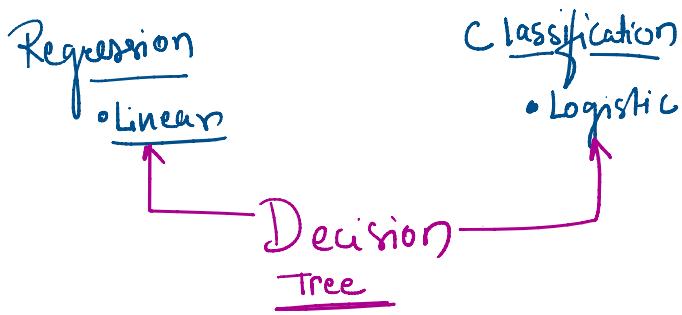


- Tree Models
 - Decision Tree



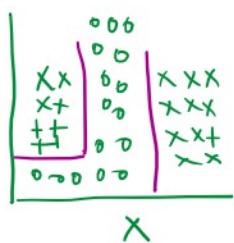
- ## Tree
- Decision trees



- Logistic Regression Limitation

• Logistic Regression (Interpretation)

①



: Linear data



②

Linear
Logistic

Scaling

Bed	Sqft	Price
3	700	
10	1500	
14	3000	
		25000
1-15	250	25000

$$y = mx + c$$

③

Interpretable

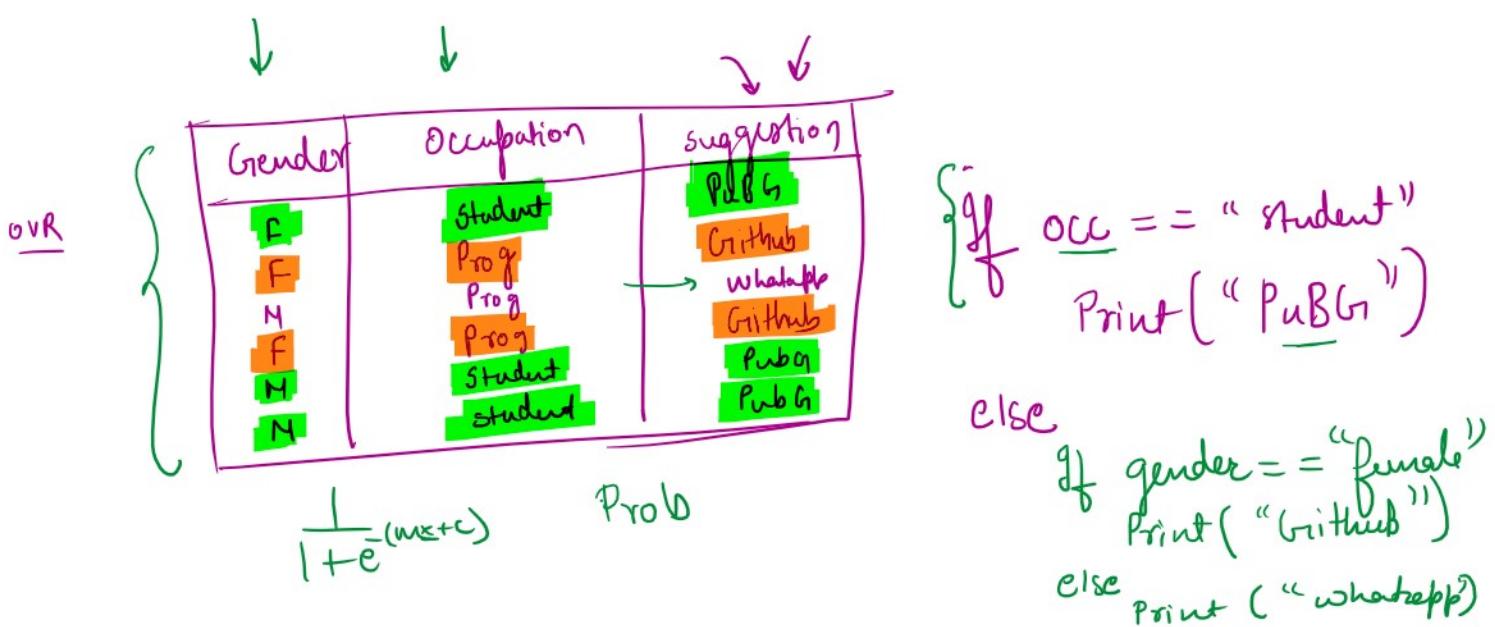
example

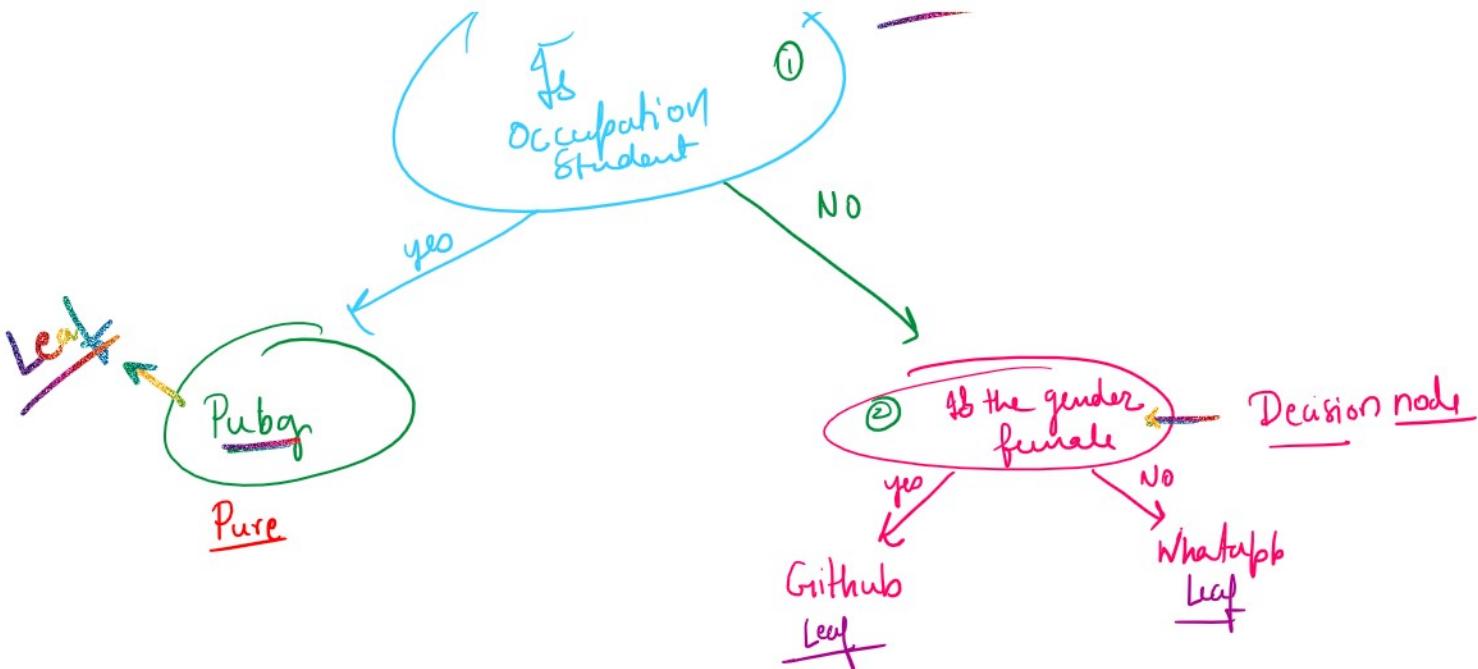
④

Important features

⑤

Complex Relationship

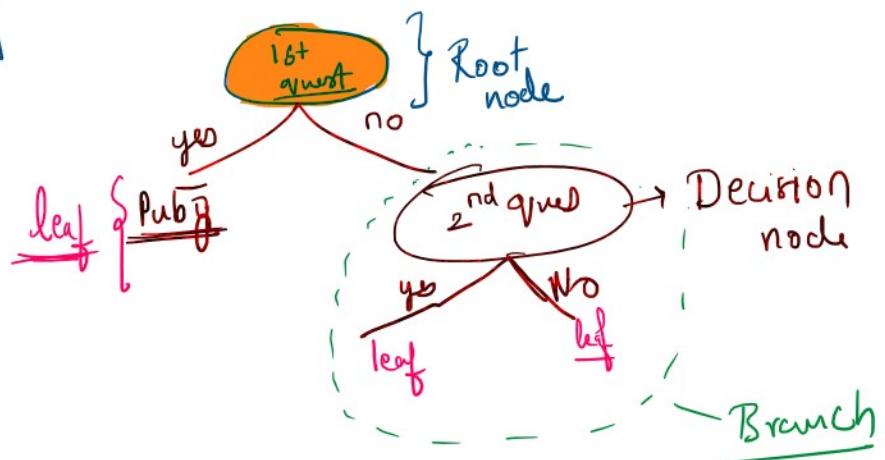




4 Decision Tree are nothing but a giant structure of nested if else statement

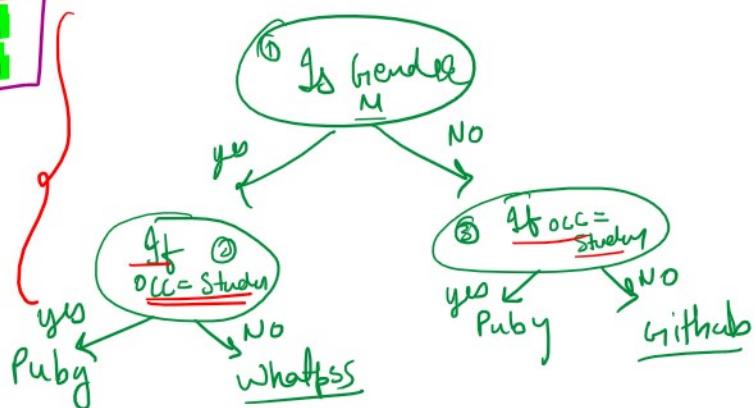
Q4:

Tree → Root
→ Branches
→ leaves



How to decide the first question

Gender	Occupation	suggestion
F	Student	PubG
F	Prog	Github
M	Prog	WhatsApp
F	Prog	Github
M	Student	PubG
M	student	PubG



Doctor

→ ① Did you eat outside Junk food } ← yes → Pure
 { ② Did you sleep well

Top question
 → That gives you max pure
 Decision

Gini Impurity

$$1 - \sum \left[p(y_i)^2 \right]$$

100 ↘ $\frac{20}{80}$ smokers
 { 80 non smokers

$$1 - \sum \left[\left(\frac{20}{100} \right)^2 + \left(\frac{80}{100} \right)^2 \right]$$

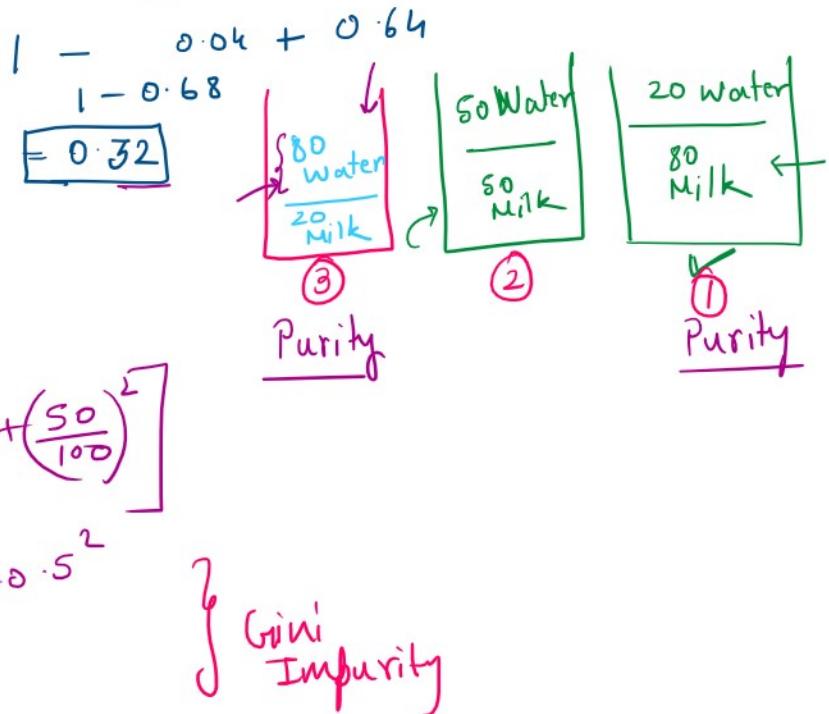
$$1 - 0.04 + 0.64$$

100 ← 50 smokers
 50 Non smokers

$$= 1 - \left[\left(\frac{50}{100} \right)^2 + \left(\frac{50}{100} \right)^2 \right]$$

$$= 1 - 0.5^2 + 0.5^2$$

$$= \boxed{0.5}$$



③ 100 ← 100 Non smoker
 0 Smoker

$$1 - \left[\left(\frac{100}{100} \right)^2 + \left(\frac{0}{100} \right)^2 \right]$$

$$1 - [1 + 0]$$

$$1 - 1$$

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

① Gini Impurity of total dataset

② Root: Calculate G.I for every column

$$\text{Cinema} - 6 - 1 - \left[\left(\frac{6}{10} \right)^2 + \left(\frac{2}{10} \right)^2 + \left(\frac{1}{10} \right)^2 + \left(\frac{1}{10} \right)^2 \right]$$

Cinema - 6
 Tennis - 2
 Stay in - 1
 Shopping - 1

$$1 - \sum \left[\left(\frac{6}{10} \right)^2 + \left(\frac{2}{10} \right)^2 + \left(\frac{1}{10} \right)^2 + \left(\frac{1}{10} \right)^2 \right] = 0.58$$

① Give Impurity for Money on the decision

Rich - 7
 Poor - 3

$$1 - \sum \left[\left(\frac{7}{10} \right)^2 + \left(\frac{3}{10} \right)^2 \right] = 0.42$$

~~Impurity in Money column~~

Rich

Cinema - 3
 Tennis - 2
 Stay in - 1
 Shopping - 1

$$1 - \sum \left[\left(\frac{3}{7} \right)^2 + \left(\frac{2}{7} \right)^2 + \left(\frac{1}{7} \right)^2 + \left(\frac{1}{7} \right)^2 \right] = 0.694$$

Poor

Cinema - 3

$$1 - \sum \left[\left(\frac{3}{3} \right)^2 \right] = 0$$

Weighted Gini

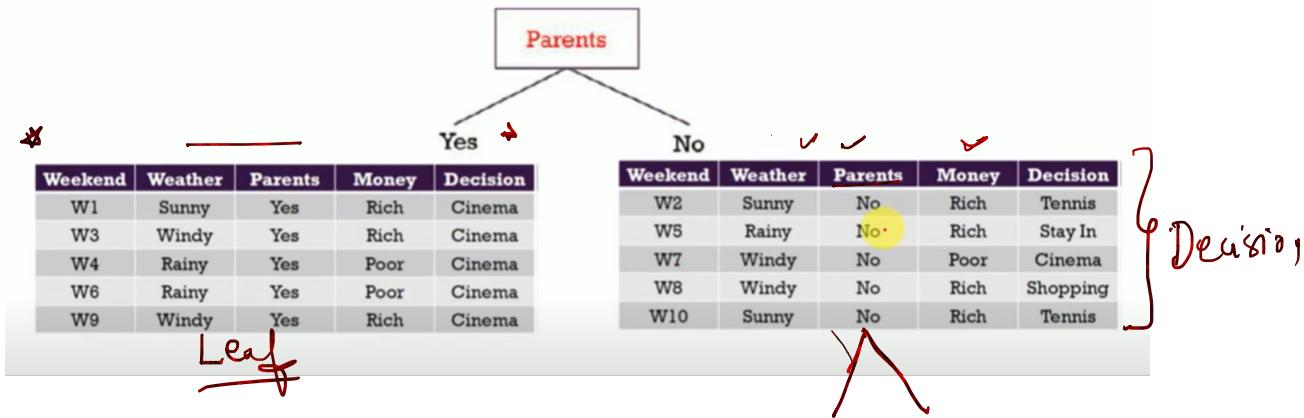
$$\boxed{\frac{7}{10} \times 0.694 + \frac{3}{10} \times 0} = 0.486$$

Money

Weather = 0.41

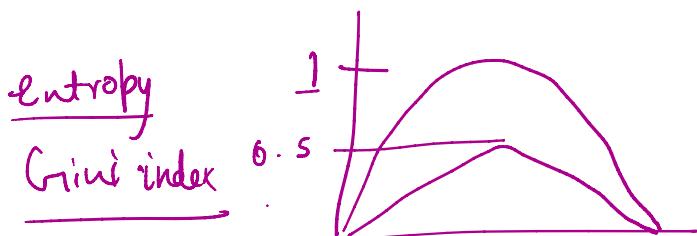
Parent = 0.36

$$\text{Money} = 0.48$$



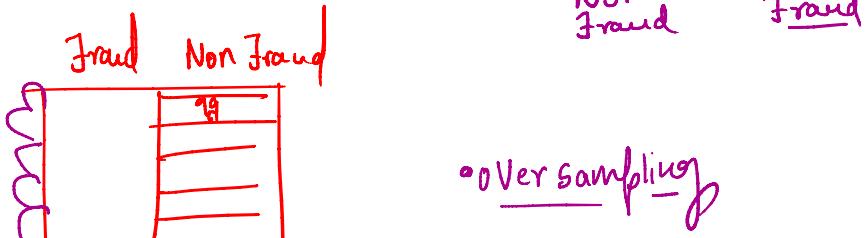
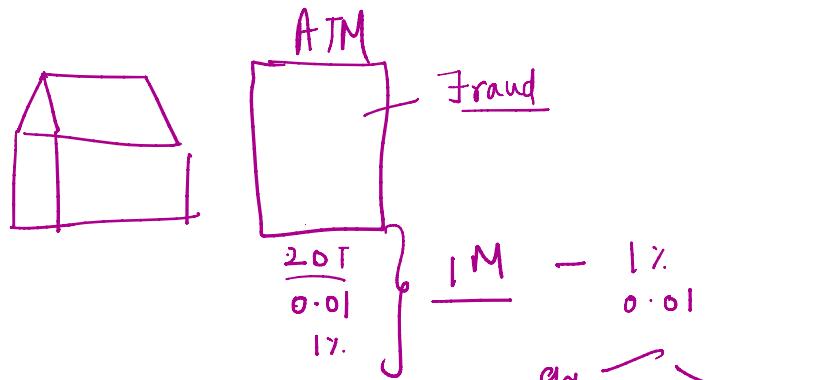
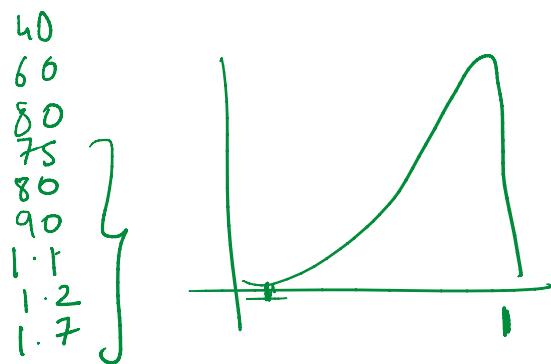
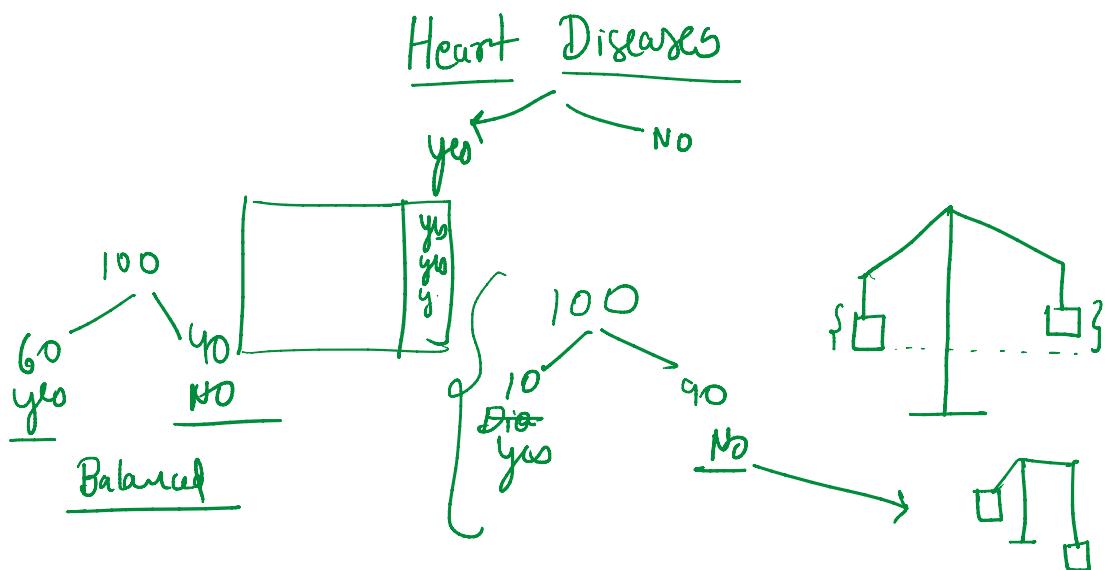
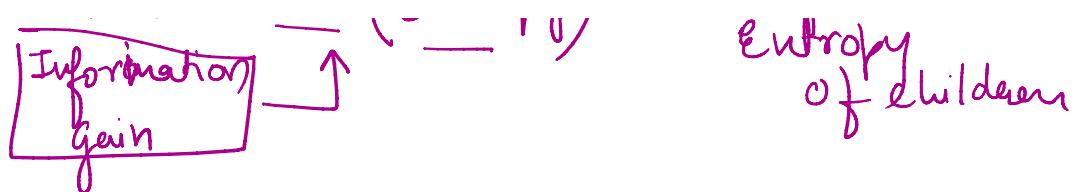
Entropy : Measure of Randomness

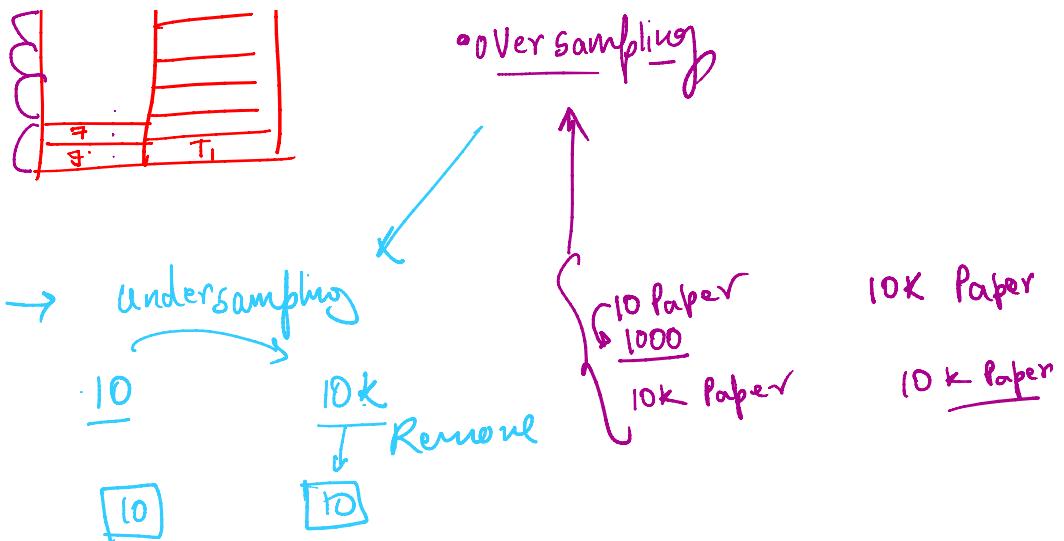
$$\begin{aligned}
 & \left(\begin{array}{c} \% \\ \% \\ \% \\ \% \end{array} \right) : -[p \log_2 p + (1-p) \log_2 (1-p)] \\
 &= -\left[\frac{3}{6} \log_2 \frac{3}{6} + \left(\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) \right] \\
 &= -\boxed{+1}
 \end{aligned}$$



IG = ① Parent (Entropy) - Weighted Entropy of children

Information ↑





SMOTE

