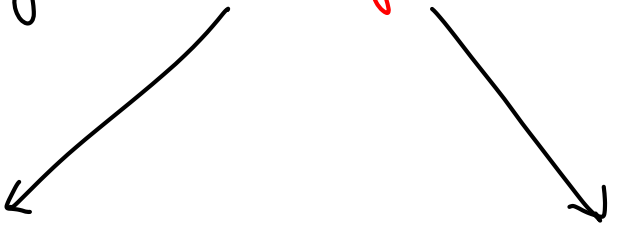
DECISION TREES

Churn ⇒ Jio, Airtel → Employee

(1) Retaining :- ↑ Salary, promotion

(2) Hiring :- training time + Cost + Misfit



Chance of an employee leaving.

Classification · Solved

Identify those important factors which are leading to employee attrition.

Solved : Model interpret & feature Importance

DS

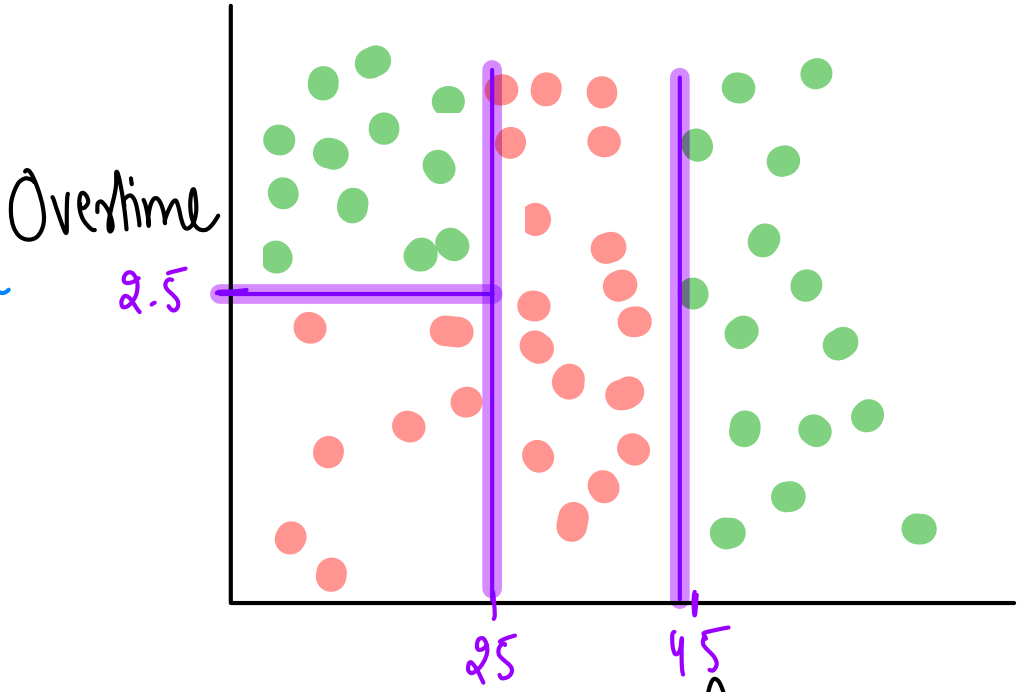
1. Data Ingestion
2. EDA
   - → preprocess / clean ✓ ✓
   - → plot distribution
   - → Encoding
   - → Imbalance
   - → Scaling
   - → Outliers
   - → Dimensionality Reduction [if required]
   - → feature engineering

⇒ Stays

⇒ leaves.

log Reg ⇒ lead to a complex model ⇒ Overfitting

KNN ⇒ Slow



if age > 45 ⇒ leave

else if age > 25 ⇒ stay

else if overtime > 2.5 = leaves

nested if else conditions

else ⇒ stays.

Axis parallel Decision Boundaries.

if age < 25

if overtime > 2.5

⇒ leave

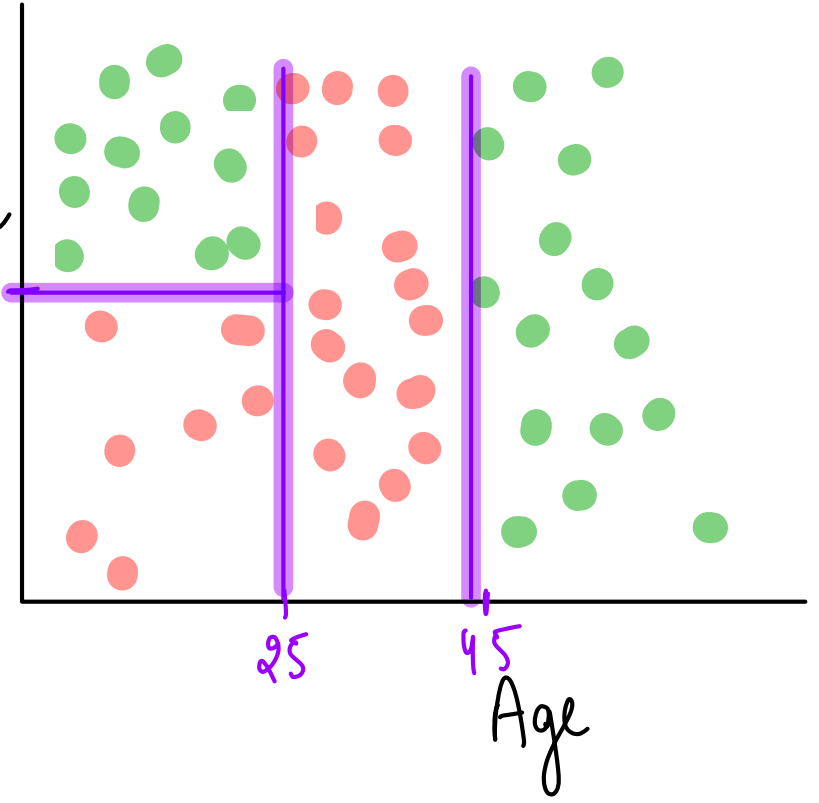
else ⇒ stay

else if age > 45

⇒ leave

else

⇒ stay.



if age < 25

if overtime > 2.5

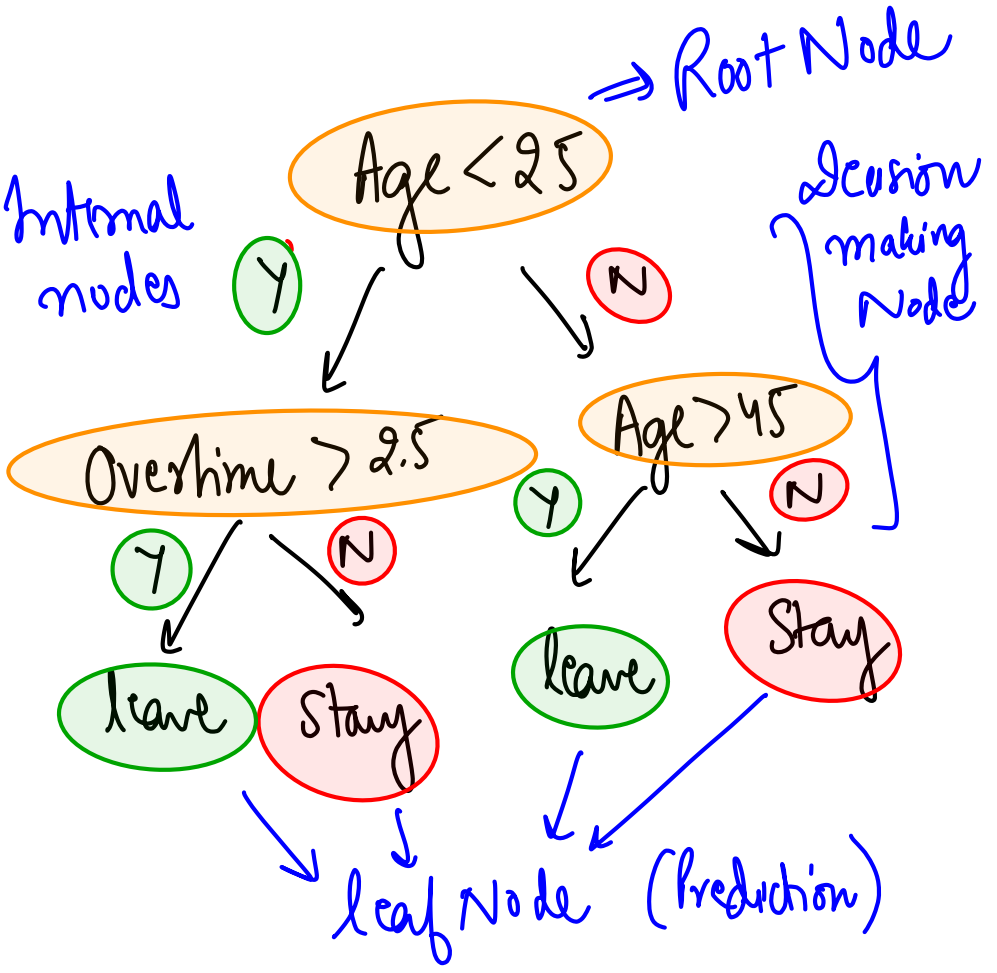
⇒ leave

else ⇒ stay

else if age > 45

⇒ leave

else ⇒ stay.



Quiz time!

Time Left: 7s

Quiz - 1 Try it out

What is the main disadvange of KNN ?

41 users have participated

✓	A	It becomes computationally expensive for large datasets	83%
	B	Not suitable to handle categorical features	5%
	C	Can't be used for regression	0%
	D	Non parametric algo so difficult to train.	12%

End Quiz Now

Quiz time!

Quiz Ended!

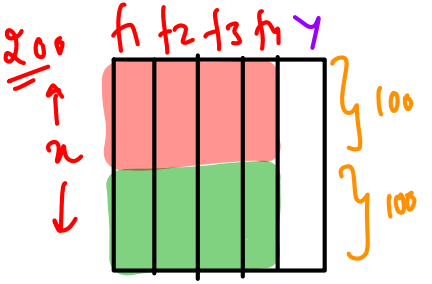
Statement I : Topmost node is called leaf node Statement II: Topmost node is called root node. Statement III: Bottom nodes are called root nodes. Statement IV: Nodes in between root and leaf are called decision nodes/ internal nodes. Which of the following statements are true ?

40 users have participated

A	All statements are true	5%
B	II & IV	83%
C	I & III	2%
D	III & IV	10%

To learn

Data: 200 $y^+ = 100$ $y^- = 100$



(I)

$n = 200$
$y^+ = 100$
$y^- = 100$

$f_i > \tau_i$

Y N

+ve -ve

$n = 100$ $n = 100$
$y^+ = 90$ $y^+ = 10$
$y^- = 10$ $y^- = 90$

misclassified = 20

(II)

$n = 200$
$y^+ = 100$
$y^- = 100$

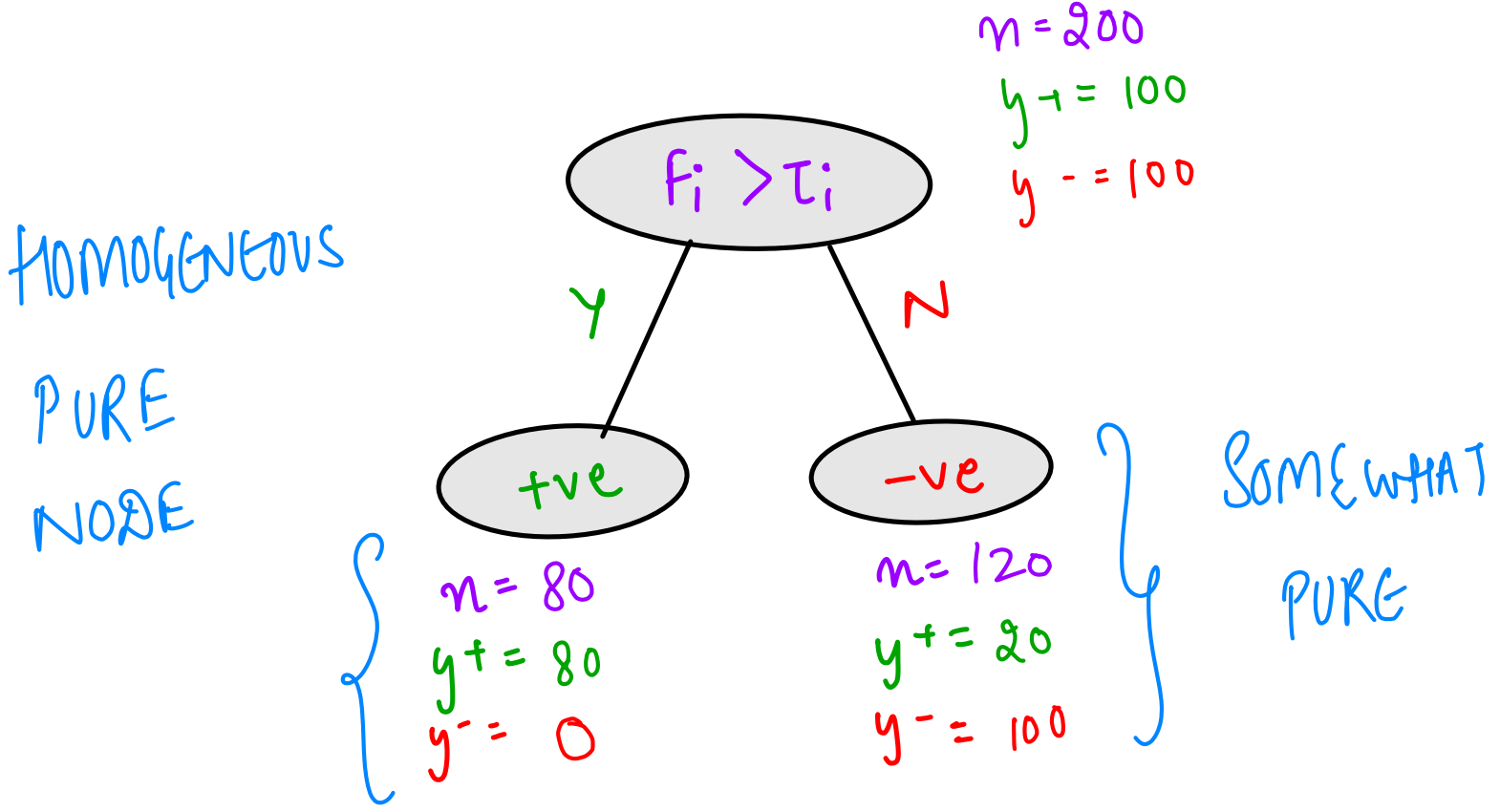
$f_j > \tau_j$

Y N

+ve -ve

$n = 100$ $n = 100$
$y^+ = 60$ $y^+ = 40$
$y^- = 40$ $y^- = 60$

misclassification = 80


n = 200
y+ = 100
y - = 100
fi > τi
Y
N
+ve
-ve
n = 80
y+ = 80
y- = 0
n = 120
y+ = 20
y- = 100
HOMOGENEOUS
PURE
NODE
SOMEWHAT
PURE

OBJECTIVE: find out those conditions which lead to creation of PURE NODES / HOMOGENEOUS REGIONS.

↳ Quantifiable measure measurement of purity of node.

Entropy

⇒ Randomness

⇒ Basics of probability

⇒ Information theory

⇒ Measure of "Impurity"

Impurity ↑ ⇒ Heterogeneity ↑ ⇒ Entropy ↑ ⇒ Purity ↓↓

Y ⇒ discrete RV $\in \{Y_1, Y_2, Y_3, Y_4\}$

$$H(Y) = -\sum_{i=1}^{k} P(y_i) \log_2 P(y_i)$$

Entropy for binary classification $\quad Y : \{0, 1\}$

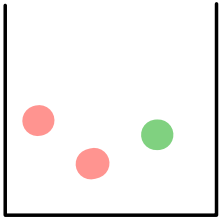
$$H(Y) = -\left[P(0) \cdot \log P(0) + P(1) \log P(1) \right]$$

$$P(0) = p \qquad P(1) = 1-p$$

$$\boxed{H(Y) = -\left[p \log p + (1-p) \log (1-p) \right]}$$

Looks similar to log loss.

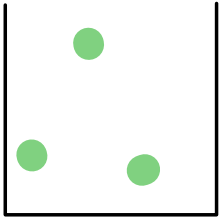
$$\text{log loss} = -\left[\underbrace{y}_{\text{actual}} \log \underbrace{\hat{y}}_{\text{predicted}} + (1-y) \log (1-\hat{y}) \right]$$



$P(R) = 1/2$
$P(G) = 1/2$

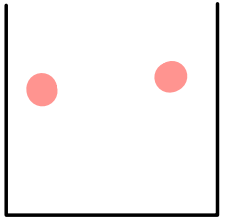
$H(Y) = 1$

$$H(Y) = -\left[P(G)\log P(G) + P(R)\log P(R) \right]$$



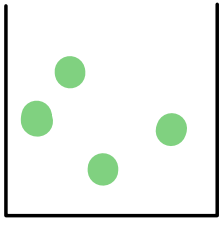
$P(R) = 1/6$
$P(G) = 5/6$

$H(Y) = 0.65$



$P(R) = 6/6 = 1$
$P(G) = 0/6 = 0$

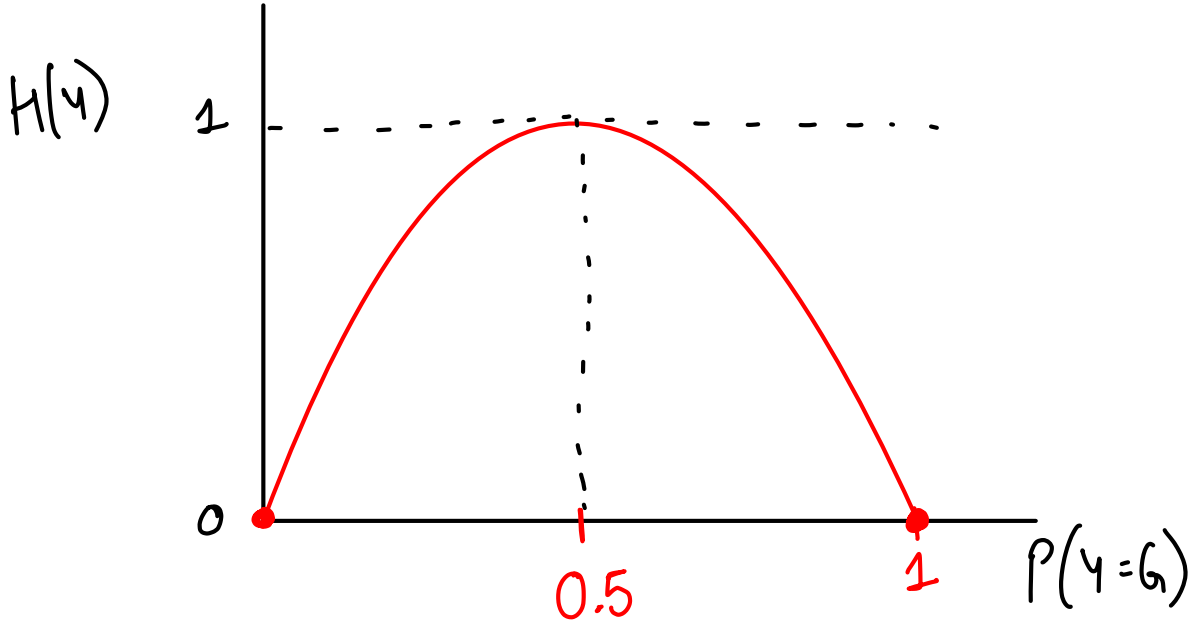
$H(Y) = 0$



$P(R) = 0/6 = 0$
$P(G) = 6/6 = 1$

$H(Y) = 0$

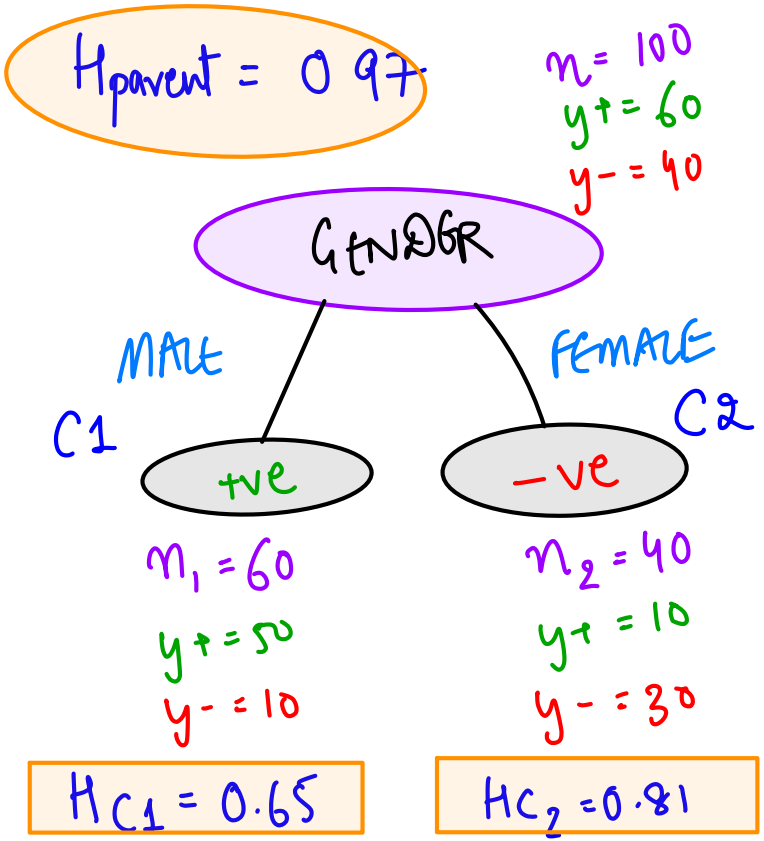
```
In [10]: def calc_entropy(p_g,p_r):
             return -((p_g*np.log2(p_g))+(p_r*np.log2(p_r)))

In [11]: calc_entropy(p_g=1/2,p_r=1/2)
Out[11]: 1.0

In [15]: calc_entropy(p_g=10e-100,p_r=1).round(2)
Out[15]: 0.0

In [16]: calc_entropy(p_g=1,p_r=10e-100).round(2)
Out[16]: 0.0

In [17]: calc_entropy(p_g=5/6,p_r=1/6)
Out[17]: 0.6500224216483541
```

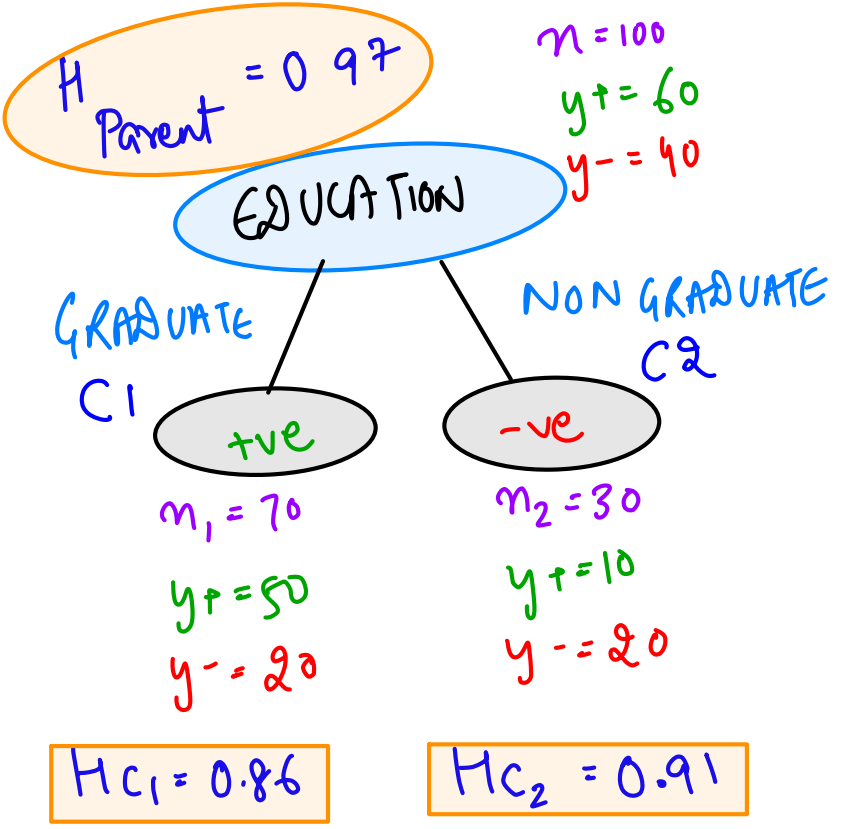

H(y)
1
0
0.5
1
P(y=G)

Decision Trees

features

Gender , Education





Information Gain

In [18]:
```
## Left
## Parent
calc_entropy(p_g=0.6,p_r=0.4)
```

Out[18]: 0.9709505944546686

In [19]:
```
## Left
## Child 1
calc_entropy(p_g=5/6,p_r=1/6)
```

Out[19]: 0.6500224216483541

In [20]:
```
## Left
## Child 2
calc_entropy(p_g=1/4,p_r=3/4)
```

Out[20]: 0.8112781244591328

In [ ]:

In [ ]:
```
## Right
## Parent
calc_entropy(p_g=0.6,p_r=0.4)
```

In [21]:
```
## Right
## Child 1
calc_entropy(p_g=5/7,p_r=2/7)
```

Out[21]: 0.863120568566631

In [22]:
```
## Right
## Child 2
calc_entropy(p_g=1/3,p_r=2/3)
```

Out[22]: 0.9182958340544896

Weighted Entropies of children.

$$\frac{n_1}{n} H(C_1) + \frac{n_2}{n} H(C_2)$$

I

↙ Gender

$H_{parent} = 0.97$

$H_{children} = 0.714$

II

↘ Education

$H_{parent} = 0.97$

$H_{children} = 0.875$

Information Gain

$$IG_I = H_{parent} - H_{children}$$ (Gender)

$$= 0.97 - 0.71 = 0.26$$

$$IG_{II} = H_{parent} - H_{children}$$ (education)

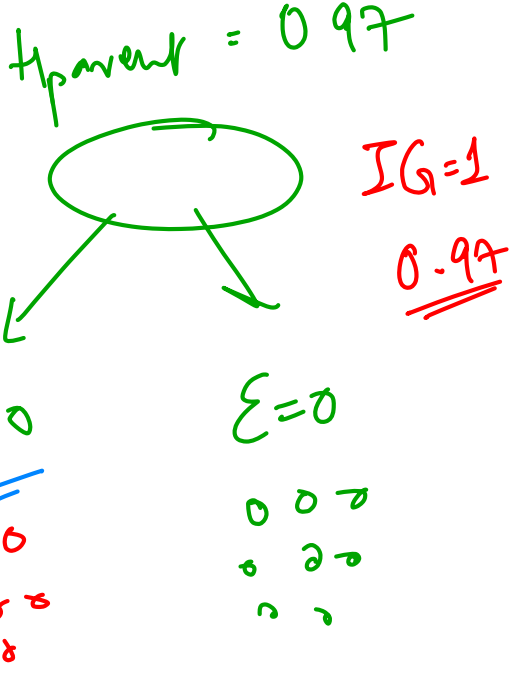
$$= 0.97 - 0.875 = 0.095$$

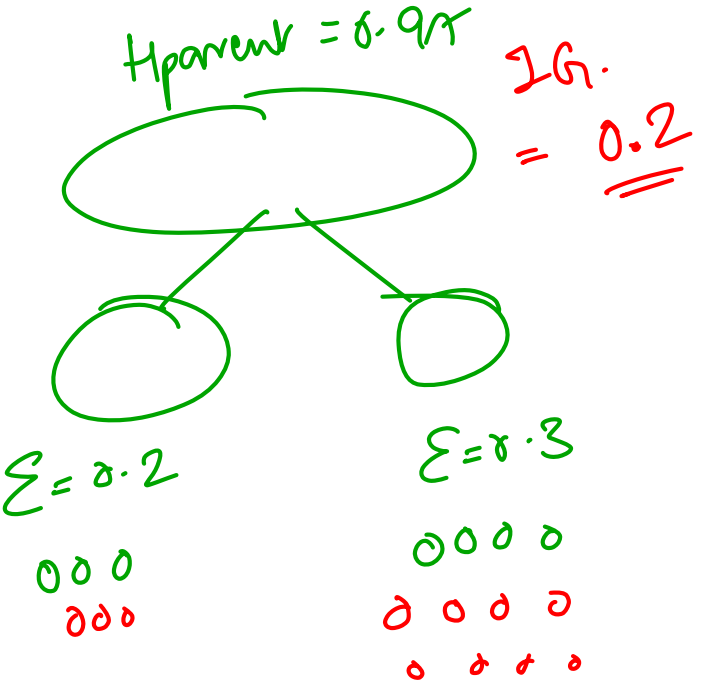
$$IG_1 > IG_2$$

We will split on Gender.

⇒ Minimise Entropy

⇒ Maximise Information Gain.





Quiz time!

Quiz Ended!

Quiz3 - Check your understanding Which of the following statement is false ?

28 users have participated

A	Purer the node, more confidence we are in our prediction	21%
B	For making prediction, DT takes majority vote of class.	25%
C	More homogenous the data is at the node, more confident we are about our prediction	25%
✓ D	None of the above	29%

Quiz time!

Quiz Ended!

Quiz5 - Check your understanding What will the value of entropy for following distribution of datapoints in node: Positive class: 50 Negative class: 0

29 users have participated

A	0.33	7%
B	1	21%
C	-1	7%
D	0	66%

+ve = 50

-ve = 0

9

20

Quiz time!

Time Left: 7s

$f_1 = 0.3$

$f_2 = 0.03$

$f_3 = 0.2$

Quiz 6 - Check your understanding We calculated information gain for 3 features which is as follows: Feature 1 : 0.3

Feature 2 : 0.03

Feature 3: 0.2 Which feature would you pick for splitting the node ?

28 users have participated

A	Feature 1	79%
B	Feature 2	11%
C	Feature 3	11%

End Quiz Now