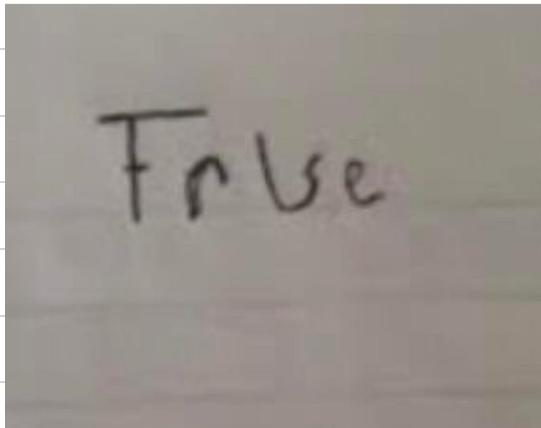
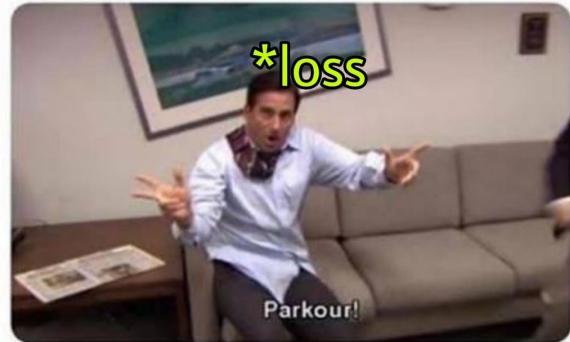


Binary classifiers be like:



When you set the learning rate little too high



Agenda

9790723608, shivam.prasad_1@scaler.com

① Decision Tree Intuition

② Nuts & Bolts

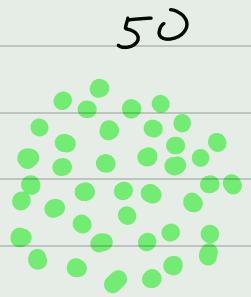
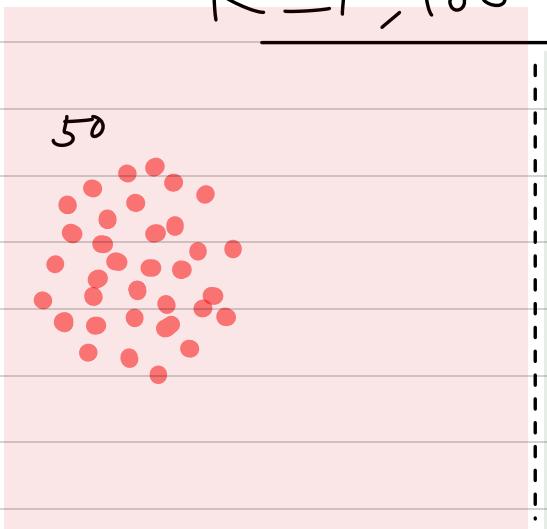
③ Entropy

④ Information Gain

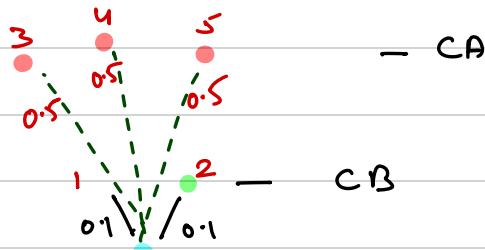


Weighted KNN

$K=1, 100$



Weighted KNN



$$\begin{aligned} w_1 &= \frac{1}{0.1} \\ &= 10 \end{aligned}$$

$$\begin{aligned} w_2 &= \frac{1}{0.1} \\ &= 10 \end{aligned}$$

$$w_3 = \frac{1}{0.5} = 2$$

$$C-G = 10 + 10 = 20$$

$$C-R = 2 + 2 + 2 = 6$$

$$w_4 = \frac{1}{0.5} = 2$$

$$w_5 = \frac{1}{0.5} = 2$$

CASE STUDY: EMPLOYEE RETENTION

HR: why should we hire you?

Me: nah, you tell me why this position is vacant? And why people leaving your company?

HR:



HR: Pls complete your 'Employee Feedback Survey'. It's completely anonymous.

Next day in office:



Quiz - 1 Try it out

What is the main disadvantage of KNN ?

0 users have participated

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

[End Quiz Now](#)

Leaderboard

Based on all quizzes from the session

	PK		Purushottam Ku...	1n 97.03
	Shreyas Gupta		Shreyas Gupta	1n 97.03
	GS		Gamidi Sri Valli S...	1n 96.70
4	Perisetta Pavan Kalyan		1n 96.67	
5	SHASHANK JHA		1n 96.56	
6	Aayush Sachan		1n 96.50	
7	Karthik		1n 96.42	
8	Narendra Babu		1n 95.60	
9	Vishwajit Verma		1n 95.37	
10	Shubham Singh		1n 93.77	



"We value you in ways
that cannot be expressed in money."



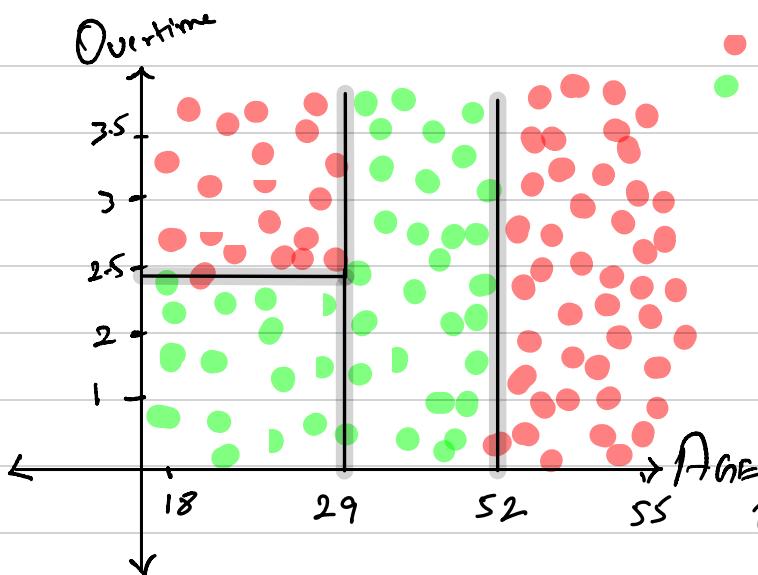
[Follow](#)

How my appraisal went this year.

8:14 AM - 5 Apr 2016

138 131

DECISION TREE - INTUITION



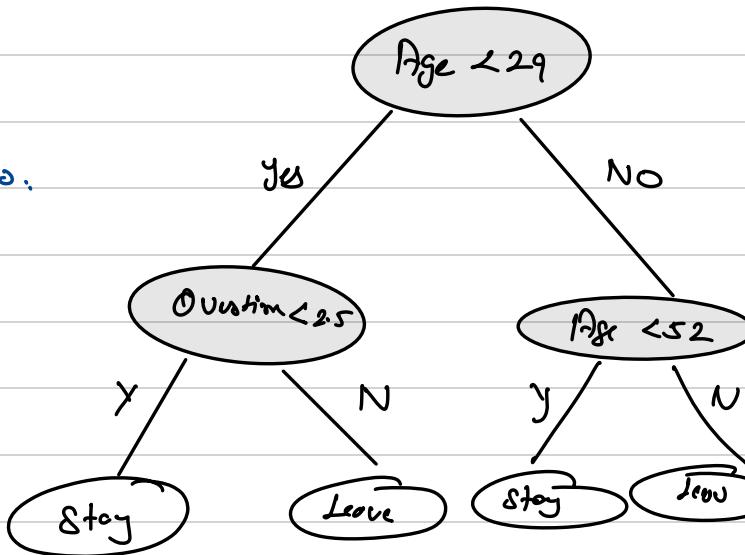
Red dot → Leave / Churn
Green dot → Stay

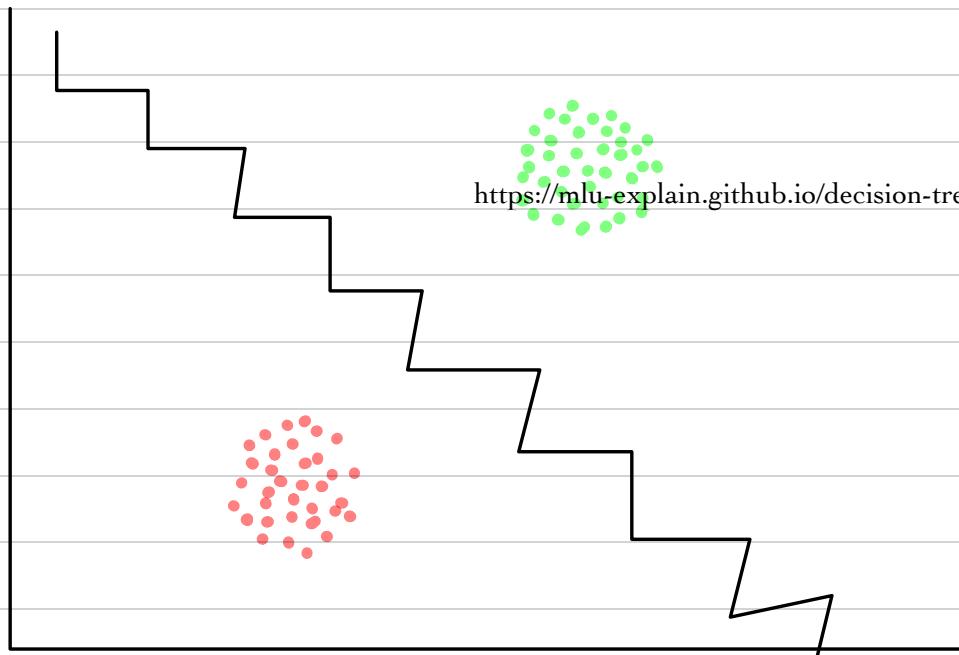
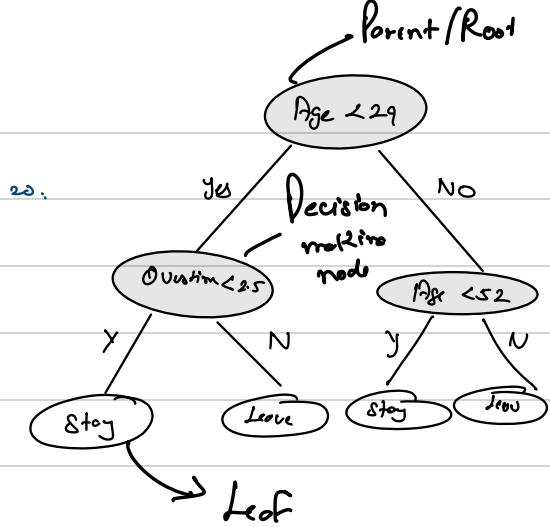
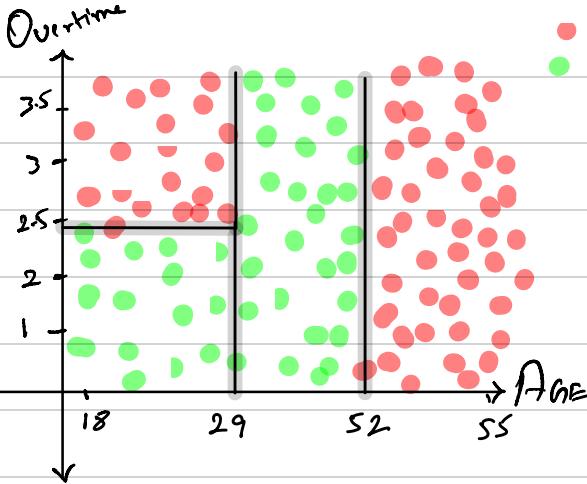
① If age > 52
Leave

② If age > 29 & age < 52
Stay

③ If age ≤ 29
If overtime < 2.5
Stay
Else if overtime > 2.5
Leave .

If age < 29:
If overtime < 2.5.
Stay
Else
Leave
Else if age < 52
Stay
Else
Leave





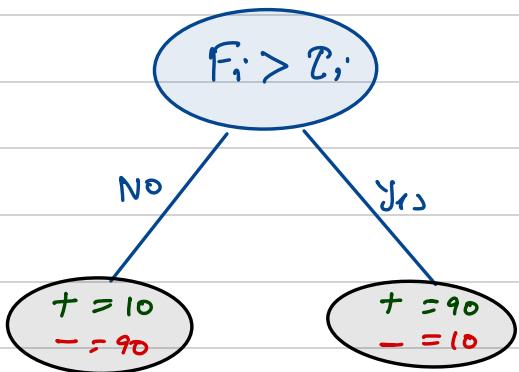
Flow To Split Nodes

Data →

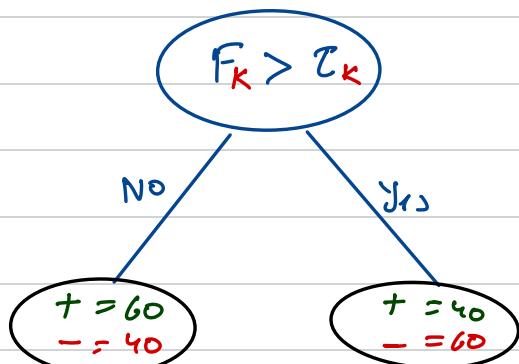
200

100 C-ve
100 (+)ve

Case - 1



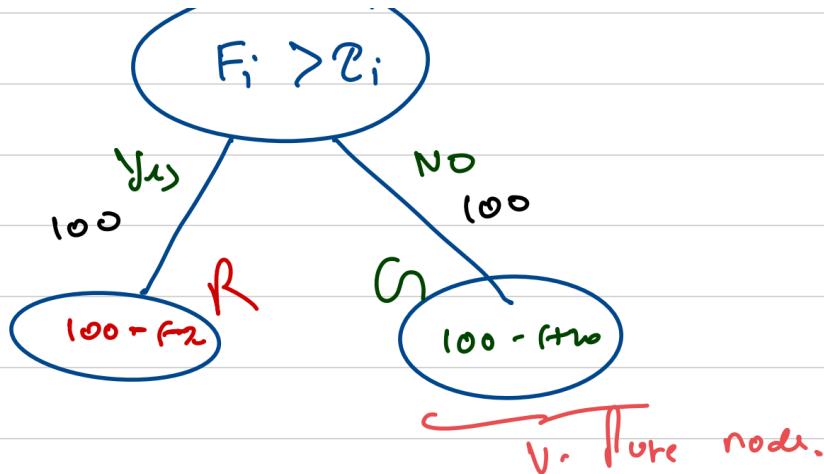
Case - 2



Alcohol addict or not

Q.1. Do you like Holdiyan bhyaja? → ^{so-A}
so-DA

Q.2: Do you like smoking? → ^{90-A}
→ 10-D-A



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

0 users have participated

- A Purer the node, more confidence we are in our prediction 0%
- B For making prediction, DT takes majority vote of class. 0%
- C More homogenous the data is at the node, more confident we are about our prediction 0%
- D None of the above 0%

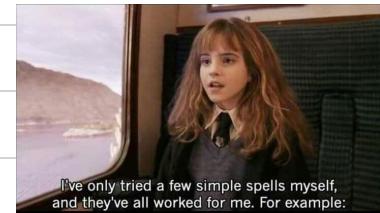
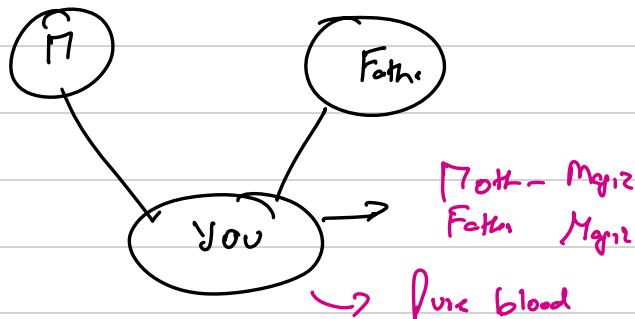
[End Quiz Now](#)

Based on all quizzes from the session

K	2	Karthik	2/3	195.00	2/3	191.20
TS	1	Tanvi Singh	3/3	226.93	2/3	188.63
AS	3	Aayush sachan	2/3	193.27	2/3	185.90
4	Pavan Kalyan	2/3	184.50	2/3	182.03	2/3
5	Gamidi Sri Valli Supraja	2/3	181.57	2/3	180.83	2/3
6	Purushottam Kumar	2/3	180.83	2/3	180.83	2/3
7	SHASHANK JHA	2/3	180.83	2/3	180.83	2/3
8	Shoreya gupta	2/3	180.83	2/3	180.83	2/3
9	Vishwajit Verma	2/3	180.83	2/3	180.83	2/3
10	Shubham Singh	2/3	180.83	2/3	180.83	2/3

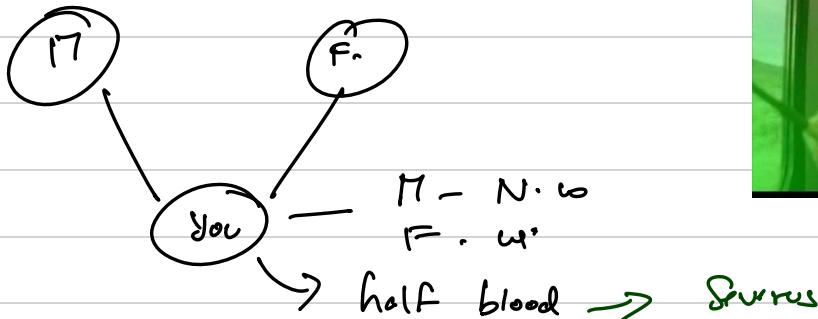
W

W₁:



N · W₁

W₁ or



N · w

N · w



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 ?

1 user has participated

- A All statements are true 0%
- B II & IV 100%
- C I & III 0%
- D III & IV 0%

[End Quiz Now](#)

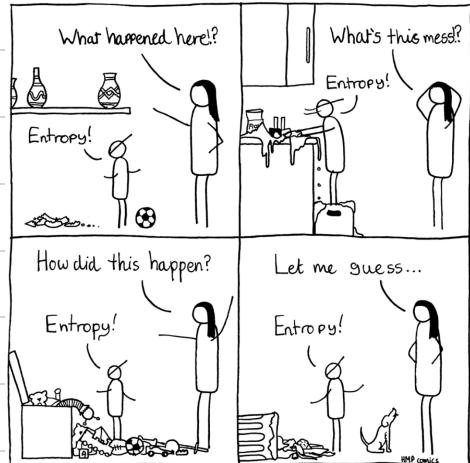
Based on all quizzes from the session

Rank	User	Score
1	AS	193.27
2	Karthik	195.00
3	PP	191.20
4	Gamidi Sri Valli Supriya	188.83
5	Purushottam Kumar	185.90
6	SHASHANK JHA	184.50
7	Shreyas gupta	182.03
8	Vishwajeet Verma	181.57
9	Shubham Singh	180.83
10	Sri Harsha Handuri	177.57

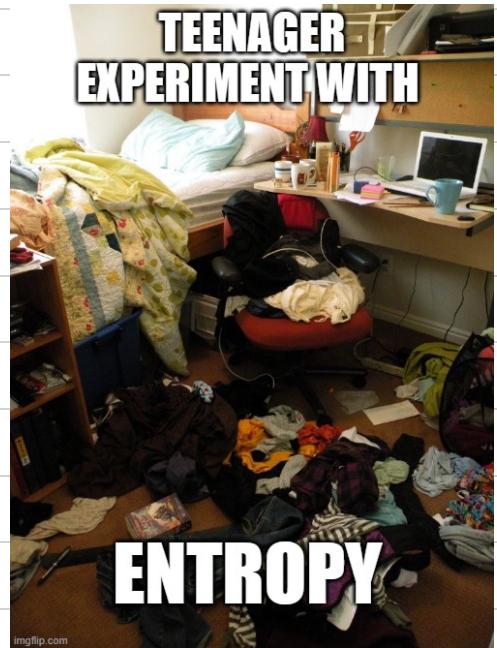
ENTROPY

* Entropy = Chaos / Degree of Randomness

$$\text{Entropy} \propto \frac{1}{\text{Purity}} \propto \frac{1}{\text{Homogeneity}}$$



This is why we don't teach our children about entropy until much later...



Entropy of K classes.

$$\hat{y} = [y_1 \quad y_2 \quad \dots \quad y_n]$$

Prob of class 1
Prob of class 2

$$y = [0, 1]$$

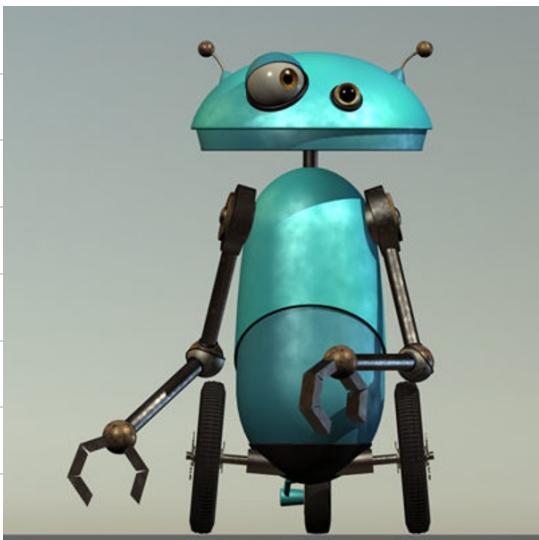
$$H(y) = - \sum_{i=1}^K P(y_i) \times \log(P(y_i))$$

$$P(1) = 1 - P(0)$$

$$H(y) = [P(0) \times \log(P(0)) + P(1) \times \log(P(1))]$$

$$= P(0) \times \log(P(0)) + (1 - P(0)) \times \log(1 - P(0))$$

Log-Loss



50% — Chars A
50% — Chars B

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

S-1

$$P(R) = 3/6$$

$$P(G) = 3/6$$

$$H(Y) = -(3/6 \times \log_2(3/6) + 3/6 \times \log_2(3/6)) = 1$$

S-2

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

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

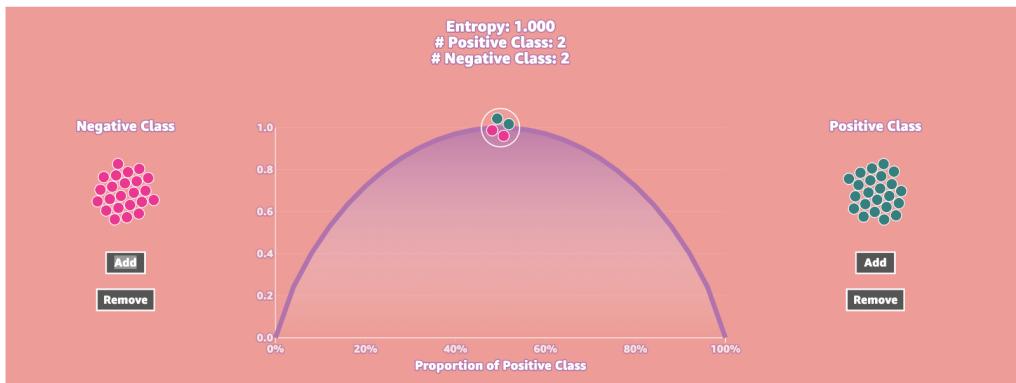
$$H(Y) = -(1/6 \times \log_2(1/6) + 5/6 \times \log_2(5/6)) = 0.65$$

S-3

$$P(R) = 0$$

$$P(G) = 1$$

$$H(Y) = -(1 \times \log_2(1)) = 0$$



Quiz4 - Try it yourself In which of the cases will the entropy be minimum ?

0 users have participated

- A A node will all datapoints belonging to one class only 0%
- B Node with datapoints belonging to both class 0%
- C Node with 100 datapoints with one datapoint belonging to positive and rest belonging to other class 0%
- D Entropy is independent of proportion of datapoints in the node. 0%

[End Quiz Now](#)

Based on all quizzes from the session



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

Negative class: 0

0 users have participated

- A 0.33 0%
- B 1 0%
- C -1 0%
- D 0 ✓ 0%

[End Quiz Now](#)

Based on all quizzes from the session

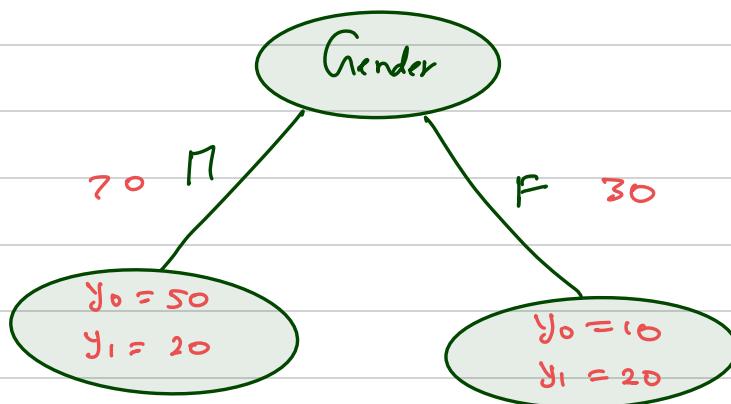


How Decision Tree Works??

$$\begin{aligned}
 n &= 100 \\
 y_1 &= 40 \\
 y_0 &= 60
 \end{aligned}$$

$$\begin{aligned}
 P(C_0) &= 60/100 = 0.6 \\
 P(C_1) &= 40/100 = 0.4 \\
 &= -(0.6 * \log_2(0.6) + 0.4 * \log_2(0.4)) = 0.971
 \end{aligned}$$

Option - 1



$$= -(50/70 * \log_2(50/70) + 20/70 * \log_2(20/70)) = 0.863$$

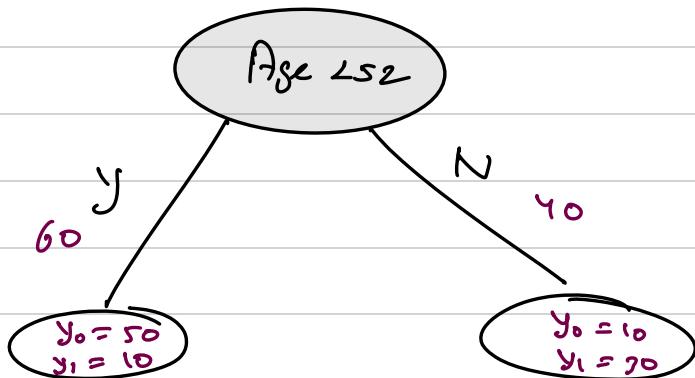
$$= -(10/30 * \log_2(10/30) + 20/30 * \log_2(20/30)) = 0.918$$

$$\left[\frac{n_1}{n_1+n_2} \times H(C_1) + \frac{n_2}{n_1+n_2} \times H(C_2) \right]$$

$$\frac{70}{100} \times 0.863 + \frac{30}{100} \times 0.918 = 0.88$$

$$0.97 - 0.88 = \underbrace{0.09}_{\text{Information gain}}$$

Option-2



$$= -(50/60 * \log_2(50/60) + 10/60 * \log_2(10/60)) = 0.65$$

$$= -(10/40 * \log_2(10/40) + 30/40 * \log_2(30/40)) = 0.811$$

$$\frac{60}{100} \times 0.65 + \frac{40}{100} \times 0.811 = 0.714$$

$$H(\text{Age}) = 0.714$$

$$\begin{aligned} \text{Information gain} &= 0.97 - 0.714 \\ &= 0.26 \end{aligned}$$

$$I_{G, \text{Gender}} = 0.09$$

$$I_{G, \text{Age}} = 0.26$$

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

Based on all quizzes from the session

Feature 2 : 0.03

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

0 users have participated

<input checked="" type="radio"/>	A	Feature 1	0%
<input type="radio"/>	B	Feature 2	0%
<input type="radio"/>	C	Feature 3	0%

[End Quiz Now](#)



Perisella Pavan ...
5/6 4 473.76



Karthik
5/6 4 478.58



Tanvi Singh
6/6 4 459.60

4	Gamidi Sri Valli Supriya	5/6 4 457.03
5	Shubham Singh	5/6 4 452.20
6	Sri Harsha Nanduri	5/6 4 442.46
7	Nachiket Pawar	5/6 4 438.73
8	Anurag Srivastava	5/6 4 399.88
9	Sumanth Andhavarapu	5/6 4 397.13
10	Purushottam Kumar	4/6 4 371.56

Which of the following is an advantage of Decision Tree

0 users have participated

A	Decision trees can handle multi class classification problems.	0%
B	Decision trees are easily interpretable	0%
C	Decision trees can capture complex nonlinear relationships in the data.	0%
D	All of the above	0%

[End Quiz Now](#)



Perisella Pavan ...
6/6 4 567.16



Karthik

6/6 4 575.06



GS

6/6 4 553.29

1	Shubham Singh	6/7 4 545.53
2	Tanvi Singh	7/7 4 542.60
3	Nachiket Pawar	6/7 4 534.36
4	Sri Harsha Nanduri	6/7 4 533.52
5	Sumanth Andhavarapu	6/7 4 491.63
6	Anurag Srivastava	6/7 4 476.88
7	Purushottam Kumar	5/7 4 465.10