NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SC1015 : Review Lecture

Classification

Dr Smitha K G

SC1015

Admin Announcements

1. Mini-Project details posted on NTULearn. Check the FAQs and take inspiration from the datasets.
2. Talk to your Lab TA to form your Project Teams. Deadline for team formation March 1st 5pm.

LAMS Completion Status

Module 1 Parts 1, 2 : Above 800 – Quiz solutions posted
Module 2 Parts 1, 2 : Above 750 – Quiz solutions posted
Module 3 : Above 600 – quiz solutions will be posted by next Monday
Module 4 : Above 250 – Complete by Exercise 5 (W7)

LAMS DS deadline: 3rd March 11.59 pm

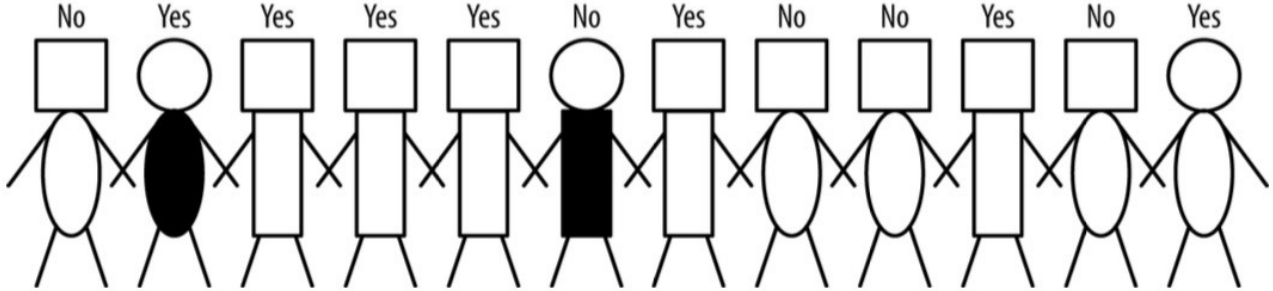
DS Theory Quiz in Recess Week : **8 March, Friday**. Slots : 12:30 pm – 2:00 pm and 2:30 pm to 4:00 pm. Lab allocations and FAQs posted.

Let’s touch upon the basic ideas of …

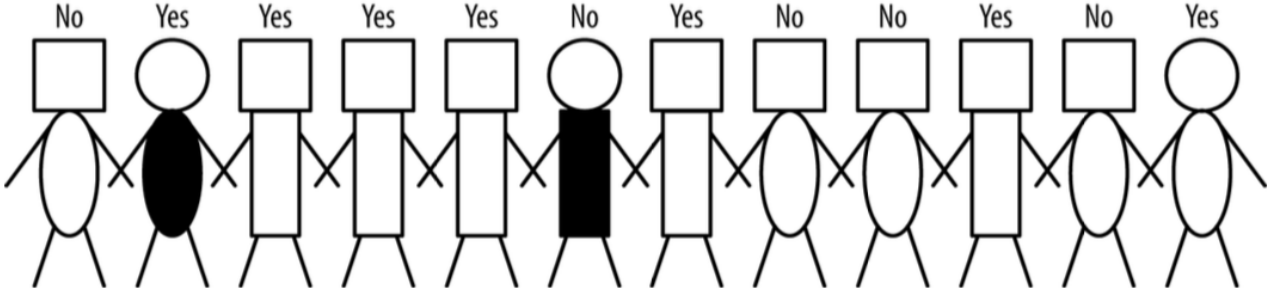
CLASSIFICATION

SC1015

Let’s clarify some of these …

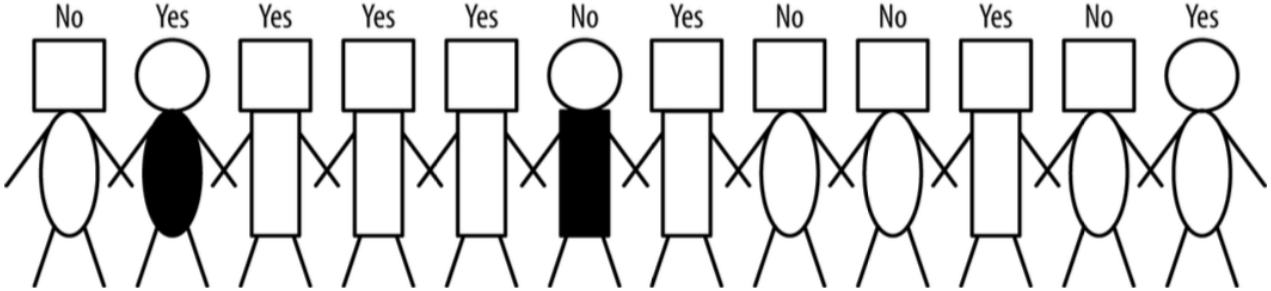


Body Shape	Body Color	Head Shape	Head Color	Funny?
Oval	White	Square	White	No
Oval	Black	Circle	White	Yes
Rectangle	White	Square	White	Yes
Rectangle	White	Square	White	Yes
Rectangle	White	Square	White	Yes
Rectangle	Black	Circle	White	No
Rectangle	White	Square	White	Yes
Oval	White	Square	White	No
Oval	White	Square	White	No
Rectangle	White	Square	White	Yes
Oval	White	Square	White	No
Oval	White	Circle	White	Yes



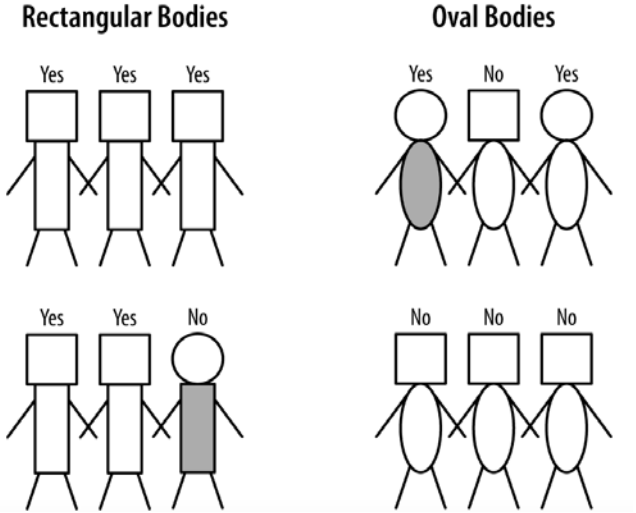
YES : NO

7 : 5



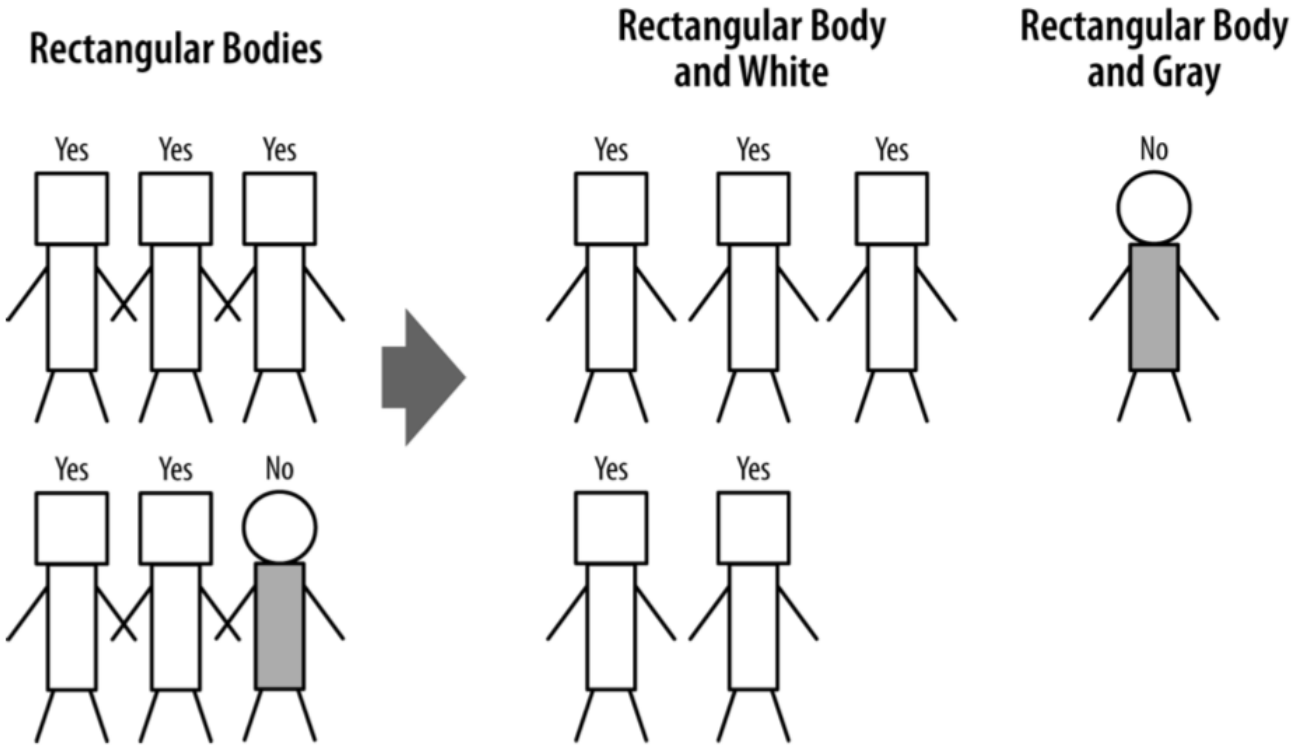
YES : NO

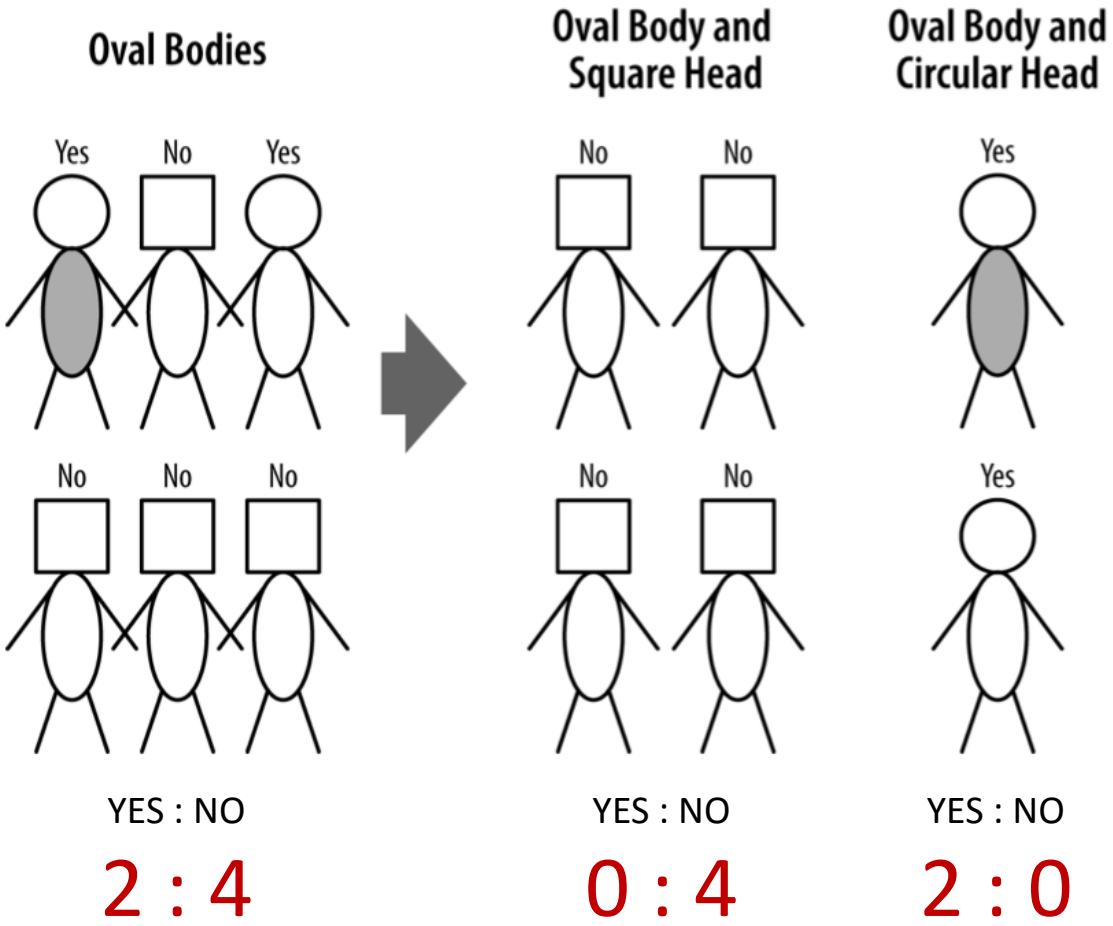
5 : 1

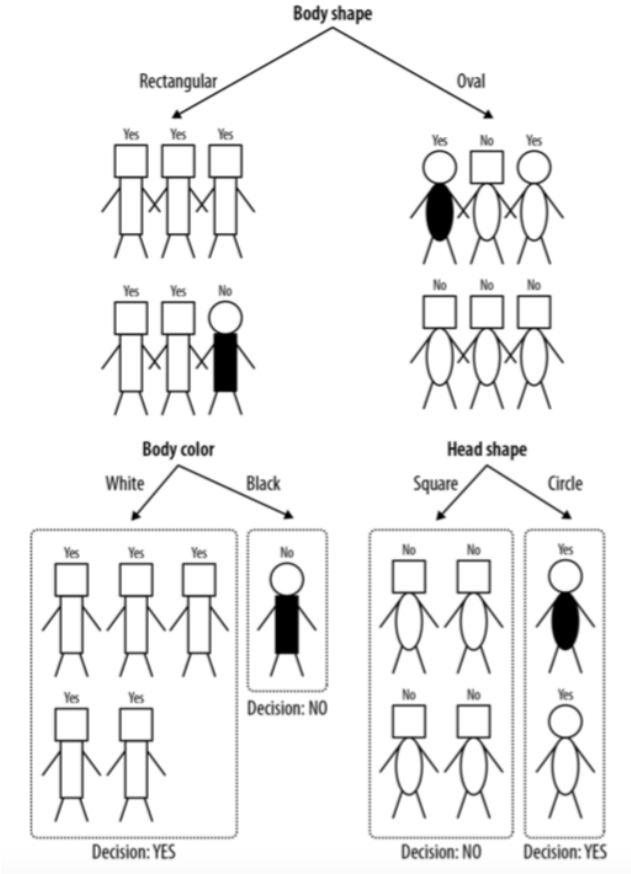


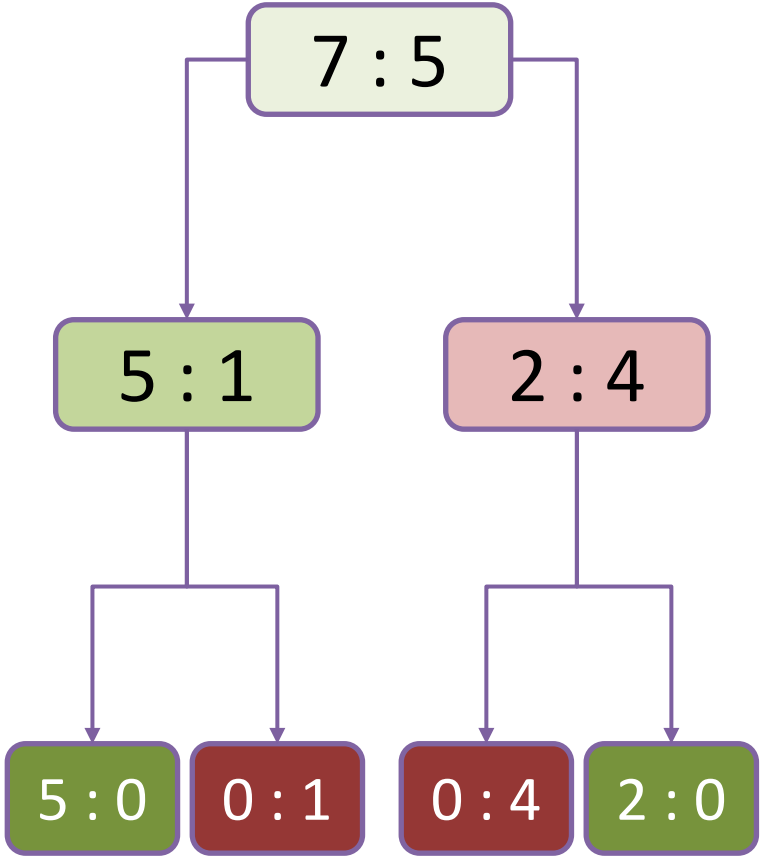
YES : NO

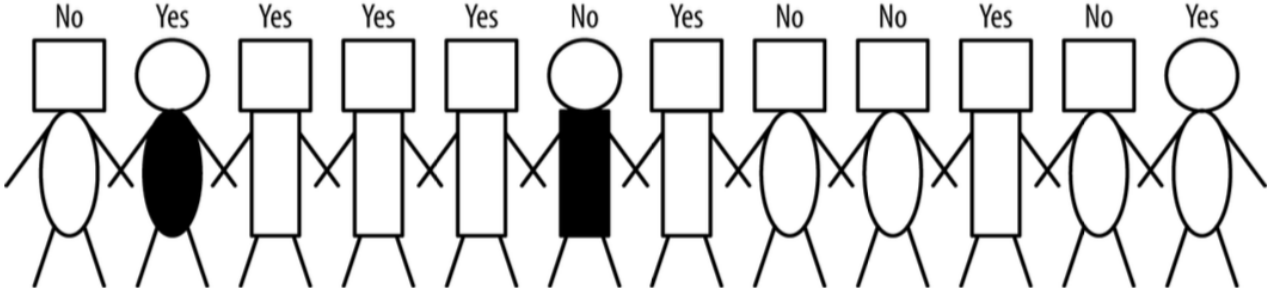
2 : 4


Rectangular Bodies
Yes
Yes
Yes
Yes
Yes
No
Rectangular Body and White
Yes
Yes
Yes
Yes
Yes
Rectangular Body and Gray
No
YES : NO
5 : 1
YES : NO
5 : 0
YES : NO
0 : 1

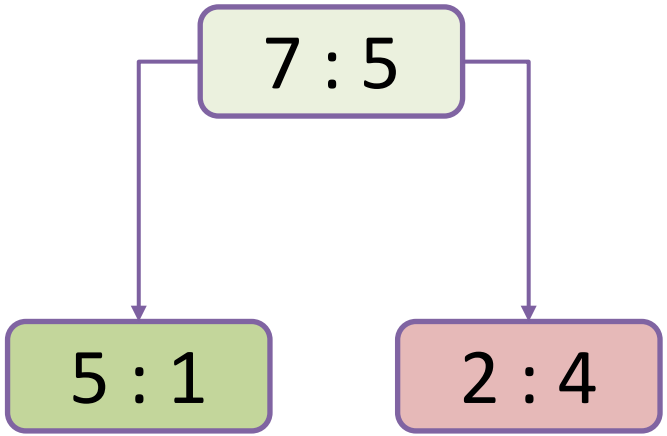

Oval Bodies
Oval Body and Square Head
Oval Body and Circular Head
Yes
No
Yes
No
No
No
No
No
No
No
No
Yes
YES : NO
2 : 4
YES : NO
0 : 4
YES : NO
2 : 0


Body shape
Rectangular
Oval
Yes Yes Yes
Yes No Yes
Yes Yes No
No No No
Body color
White
Black
Yes Yes Yes
Yes Yes
No
Decision: NO
Decision: YES
Head shape
Square
Circle
No No
No No
Yes
Yes
Decision: NO
Decision: YES

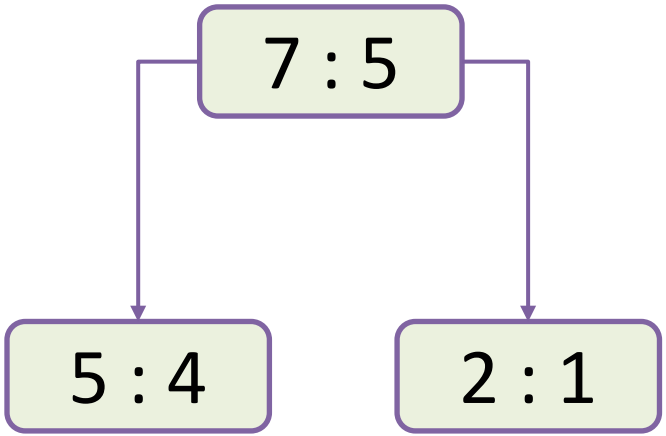

7 : 5
5 : 1
2 : 4
5 : 0
0 : 1
0 : 4
2 : 0



Body : Rectangle or Oval



Head : Square or Circle



Parent : Gini = $1 - (7/12)^2 - (5/12)^2$ = **0.486**

LChild : $1 - (5/6)^2 - (1/6)^2$ = 0.278 | 6 samples

RChild : $1 - (2/6)^2 - (4/6)^2$ = 0.444 | 6 samples

Children : 0.278 x (6/12) + 0.444 x (6/12) = **0.361**

Improvement = 0.486 – 0.361 = 0.125

Parent : Gini = $1 - (7/12)^2 - (5/12)^2$ = **0.486**

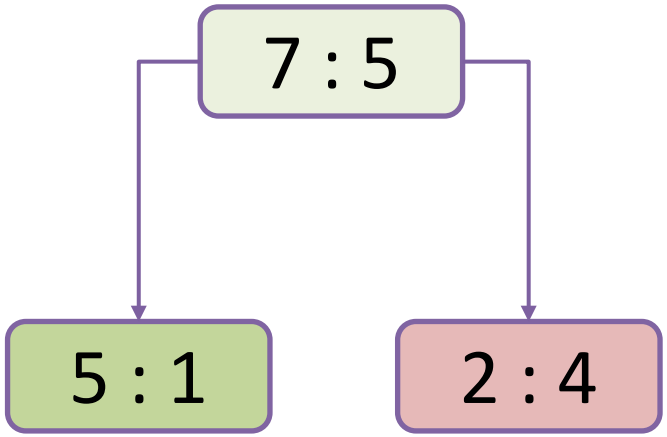
LChild : $1 - (5/9)^2 - (4/9)^2$ = 0.494 | 9 samples

RChild : $1 - (2/3)^2 - (1/3)^2$ = 0.444 | 3 samples

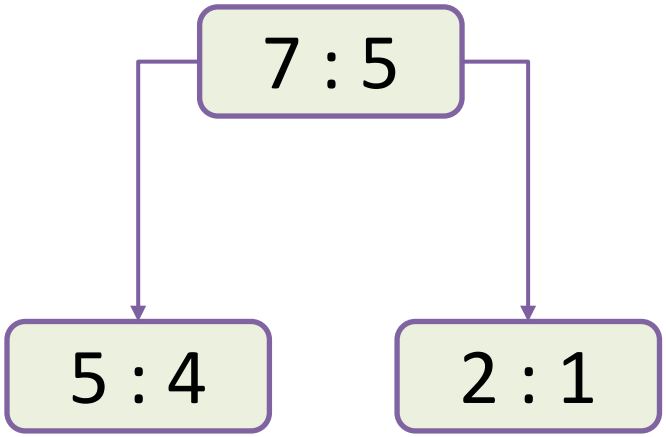
Children : 0.494 x (9/12) + 0.444 x (3/12) = **0.482**

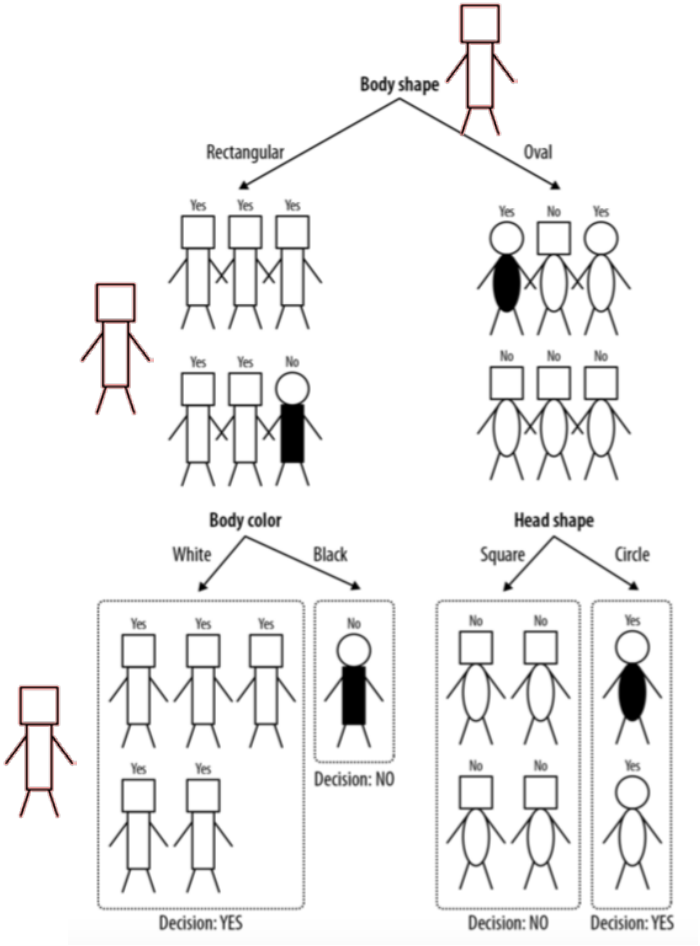
Improvement = 0.488 – 0.482 = 0.006

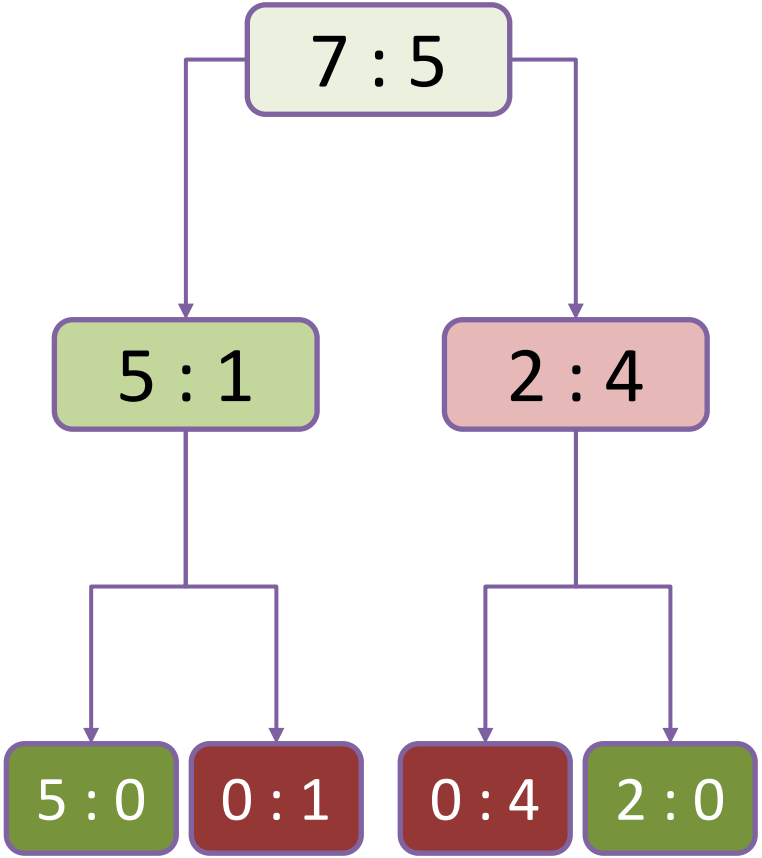
Body : Rectangle or Oval

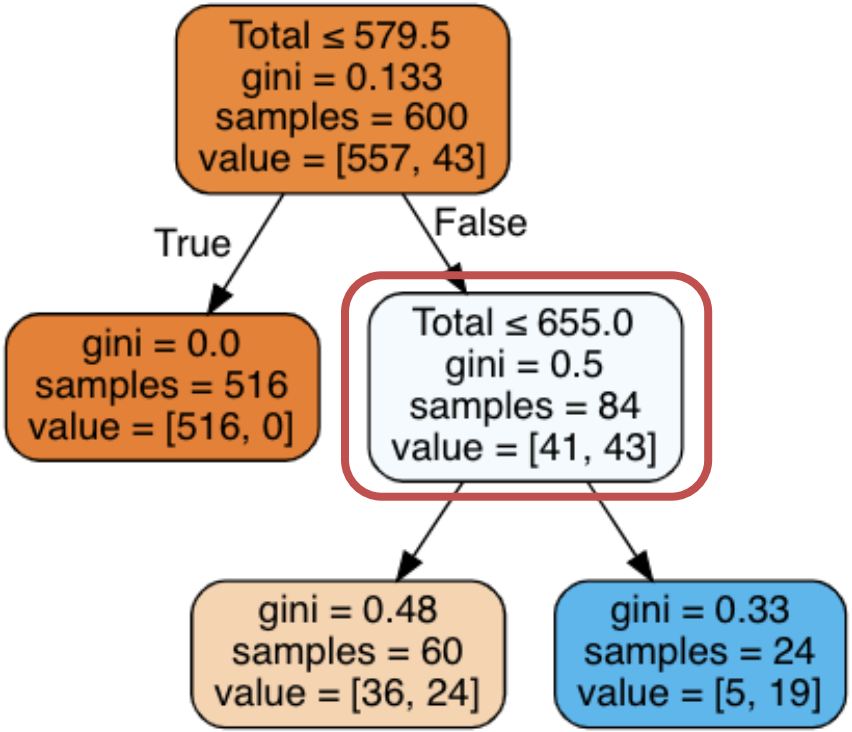


Head : Square or Circle




Body shape
Rectangular
Oval
Yes Yes Yes
Yes No Yes
Yes Yes No
No No No
Body color
White
Black
Head shape
Square
Circle
Yes Yes Yes
No
Yes Yes
Decision: NO
No No
Yes
No No
Yes
Decision: YES
Decision: NO
Decision: YES


7 : 5
5 : 1
2 : 4
5 : 0
0 : 1
0 : 4
2 : 0



Binary Classification

How to “read” a Decision Tree?

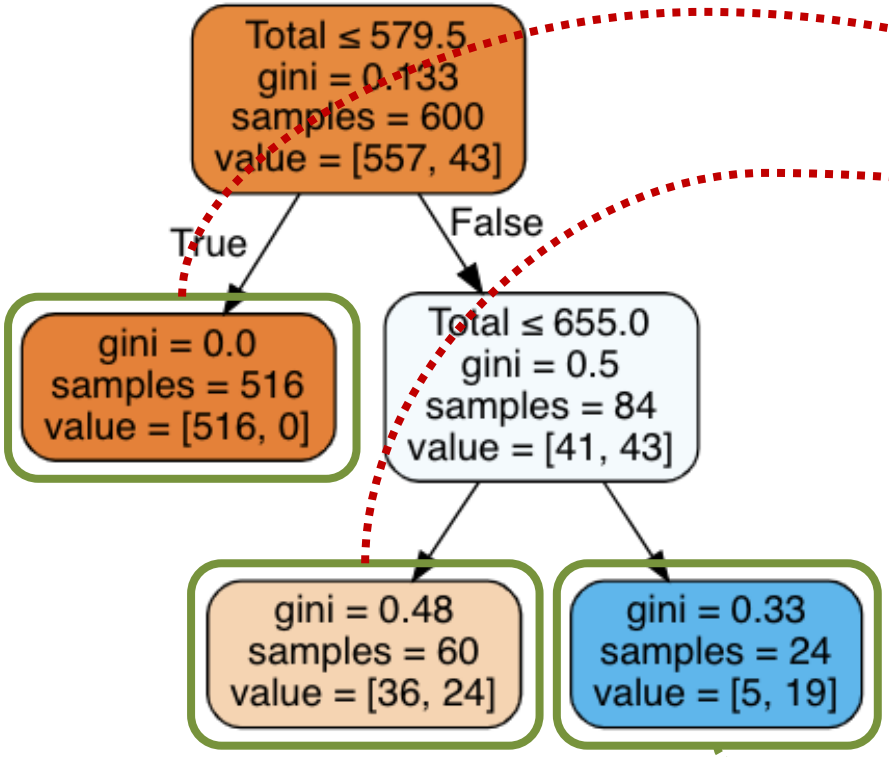
If **Total <= 655.0**, go to Left Child
Otherwise, go to Right Child

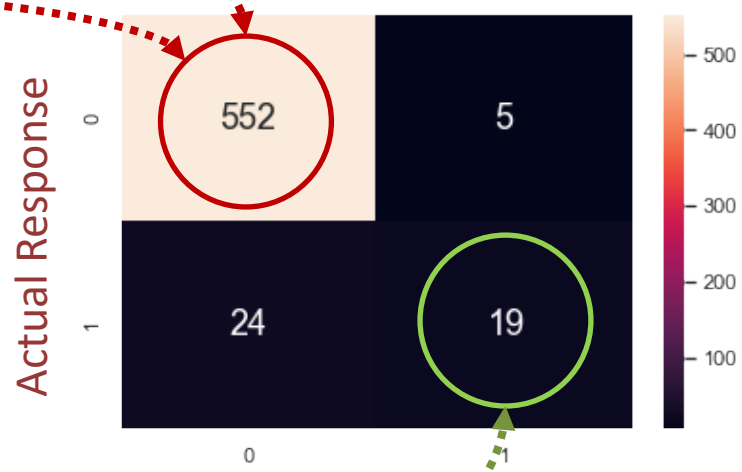
Gini = 0.5 at this node

Total samples at node = **84**
Proportion **NL : L = 41 : 43**

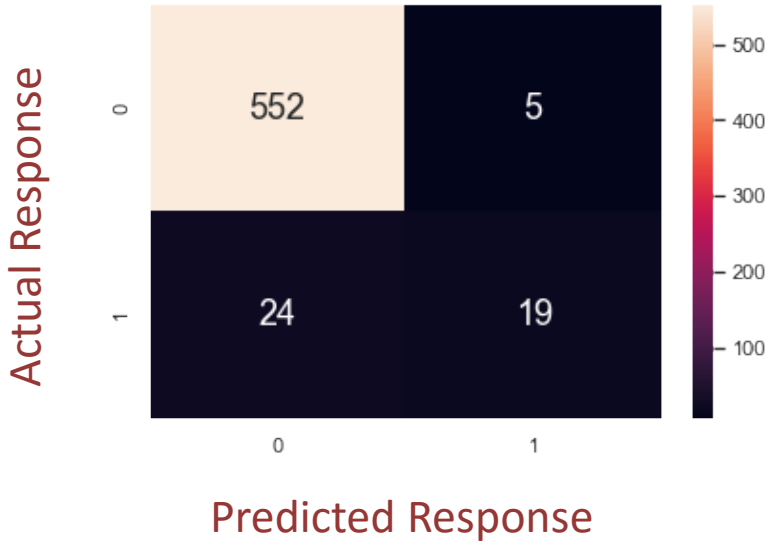
Binary Classification

Decision Tree to Confusion Matrix





Predicted Response



Data Science

Binary Classification

Goodness of Fit of the Model

TP : True predicted as True 19
TN : False predicted as False 552
FN : True predicted as False 24
FP : False predicted as True 5

$$accuracy = \frac{552 + 19}{552 + 19 + 5 + 24}$$

"Positive" : 1 Legendary
"Negative" : 0 Non-Legendary

$$tpr = \frac{19}{19 + 24}, \quad fnr = \frac{24}{24 + 19}$$

$$fpr = \frac{5}{5 + 552}, \quad tnr = \frac{552}{552 + 5}$$

Confusion Matrix : https://en.wikipedia.org/wiki/Confusion_matrix

Actual **N**	**TN**	**FP**
Actual **P**	**FN**	**TP**
	Predicted **N**	Predicted **P**

25	**0**
0	**75**

0	**25**
0	**75**

When will you be happy?

25	**0**
75	**0**

0	**25**
75	**0**

Ideal TPR = 1, FPR = 0

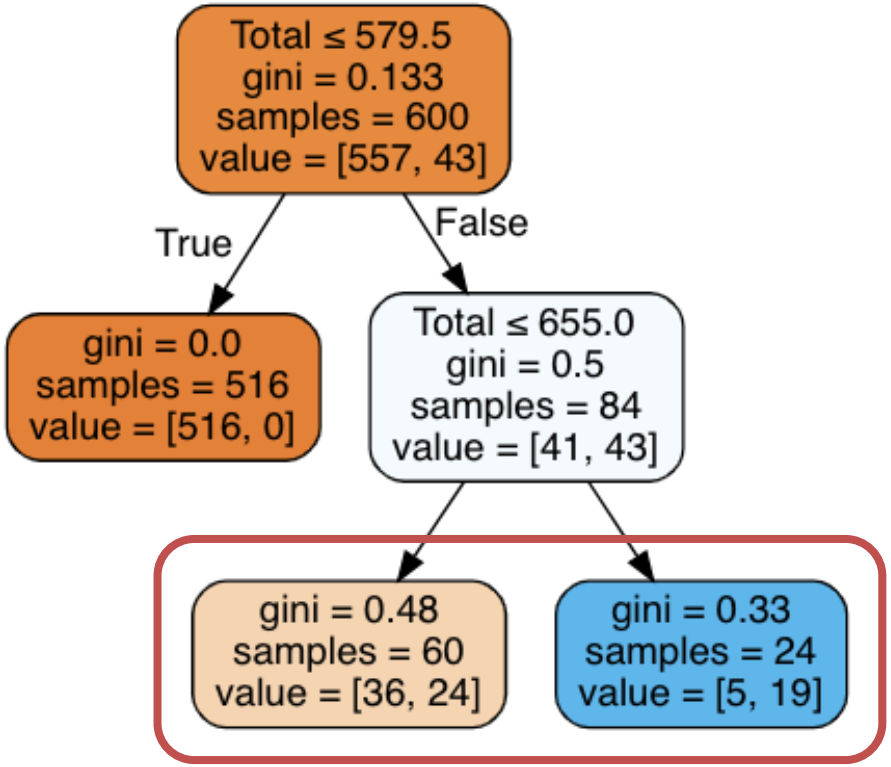
Bad? TPR = 1, FPR = 1

Bad? TPR = 0, FPR = 0

Trash TPR = 0, FPR = 1

Balancing classes to achieve the desired TPR and FPR is a tricky thing to do. ☺

F1 Score for balanced FPR and FNR : https://en.wikipedia.org/wiki/F-score



Data Science

Binary Classification

How does a Tree "decide" classes?

The tree doesn't! **You** decide it on your own by choosing **Decision Threshold**.

If Proportion > T, you call it Positive, and else, you call it Negative class.

Default Threshold for Trees = 0.5

Experiment with the Decision Threshold!

Use your tree to find Leaf Nodes.

Vary your decision threshold T in steps from 0 to 1 and note the TPR and FPR.

T = 0 : Everyone P
TPR = 1, FPR = 1

T = 1 : Everyone N
TPR = 0, FPR = 0