NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Data Science

Classification Tree

Sourav SEN GUPTA, Lecturer

School of Computer Science and Engg.
Nanyang Technological University

Second Motivation for Data Science

CLASSIFICATION IN BUSINESS

The Science

Common Problems

Prediction : Categorical

Is it Type A or Type B?

What are the chances that Sourav will default a car purchase loan?

Will DBS allow one?

https://www.youtube.com/watch?v=0XyV91VYrDs

The Science

Common Solutions

Prediction : Numeric

Regression

Prediction : Categorical

Classification

Detection : Anomalies

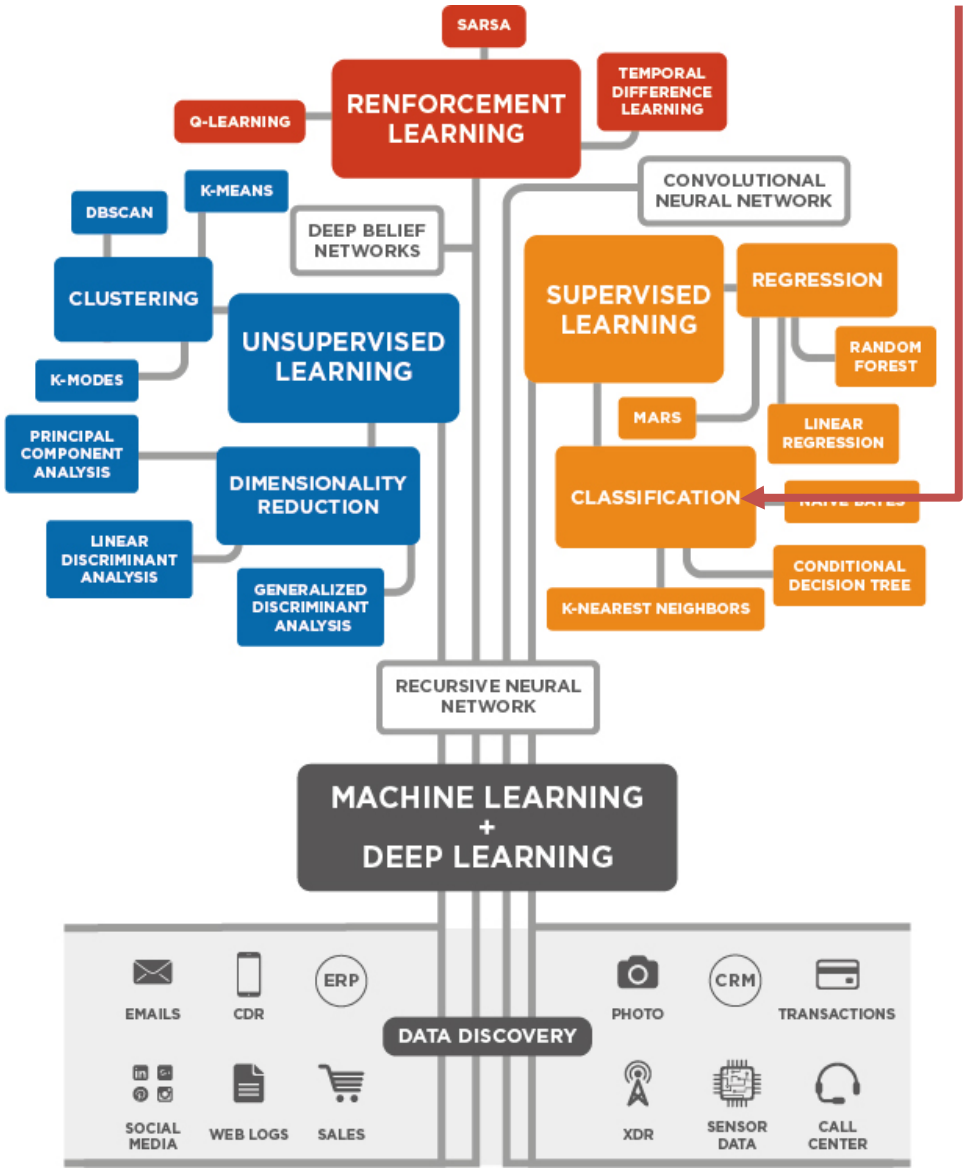
Anomaly Detection

Detection : Structure

Clustering and Dim-Red

Decision for Action

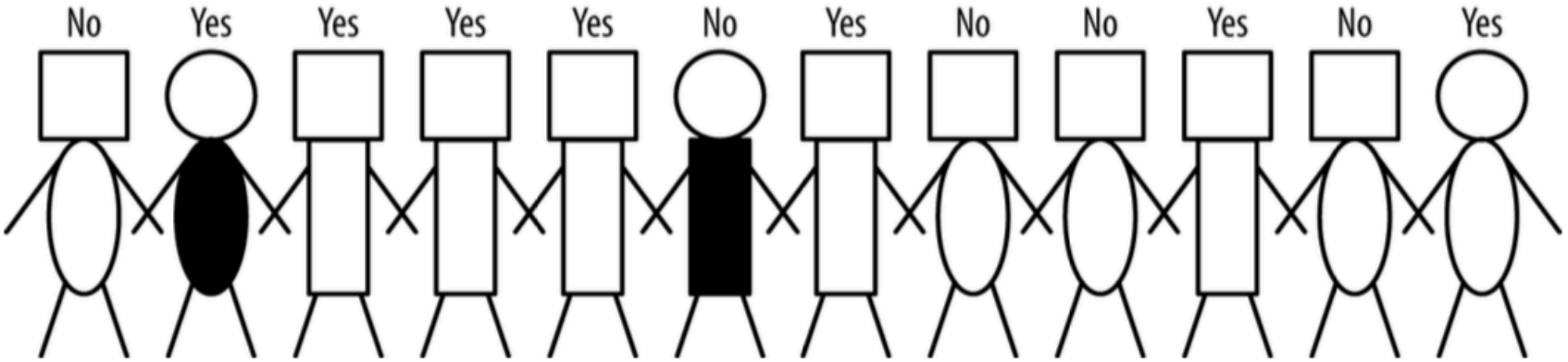
Reinforcement Learning



http://blogs.teradata.com/data-points/tree-machine-learning-algorithms/

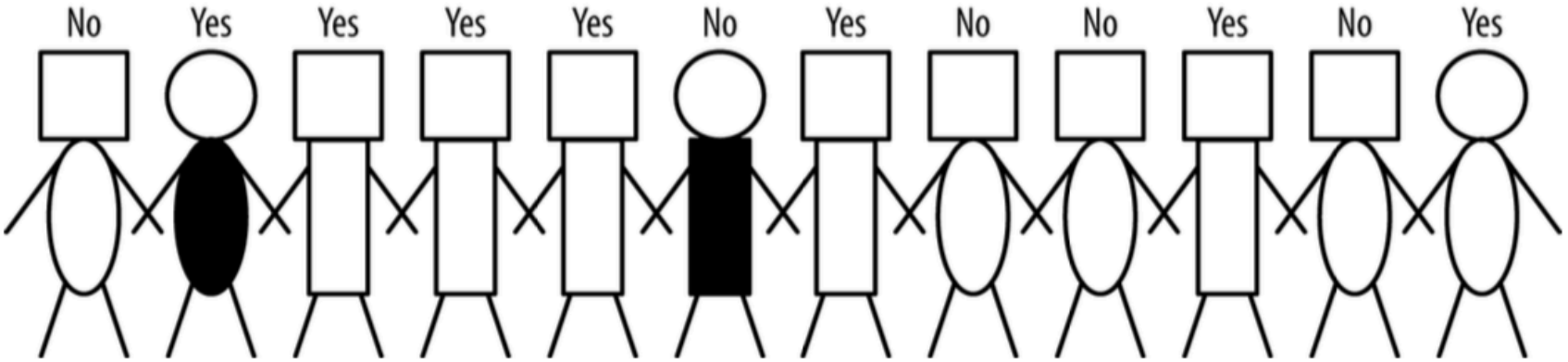
Is the credit rating “Good” or “Bad”?

HOW TO YOU DIFFERENTIATE?



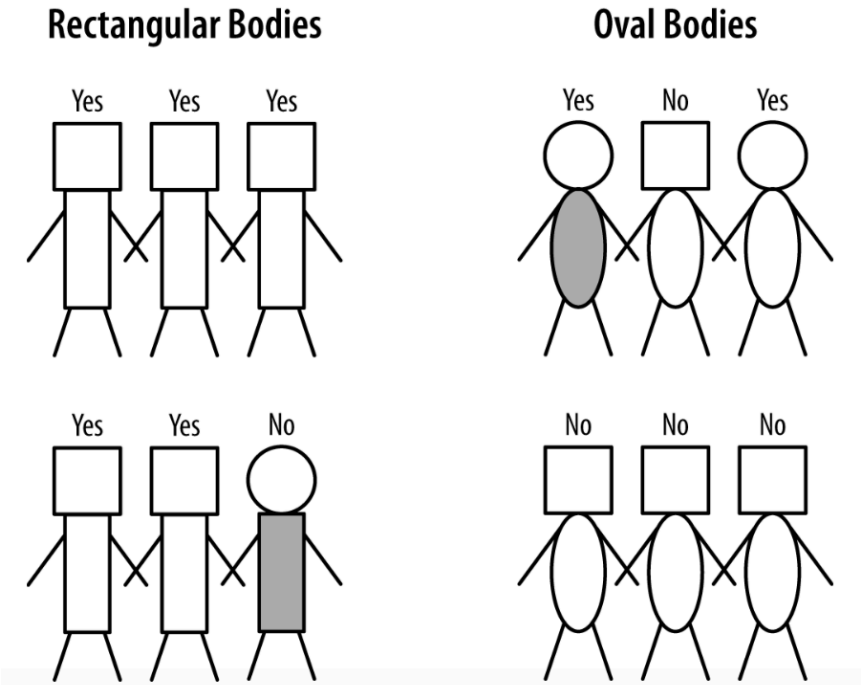
YES : NO

7 : 5



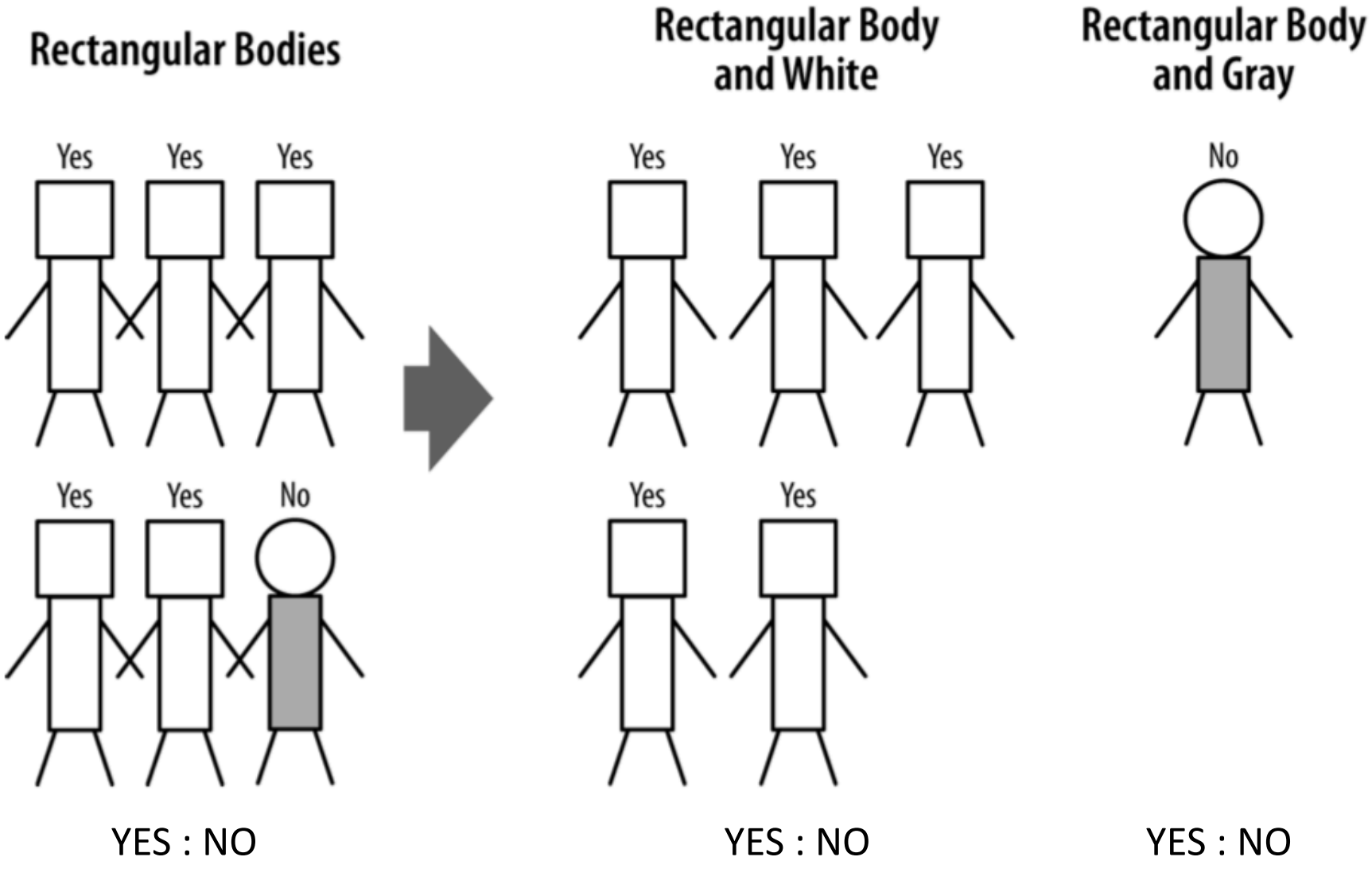
YES : NO

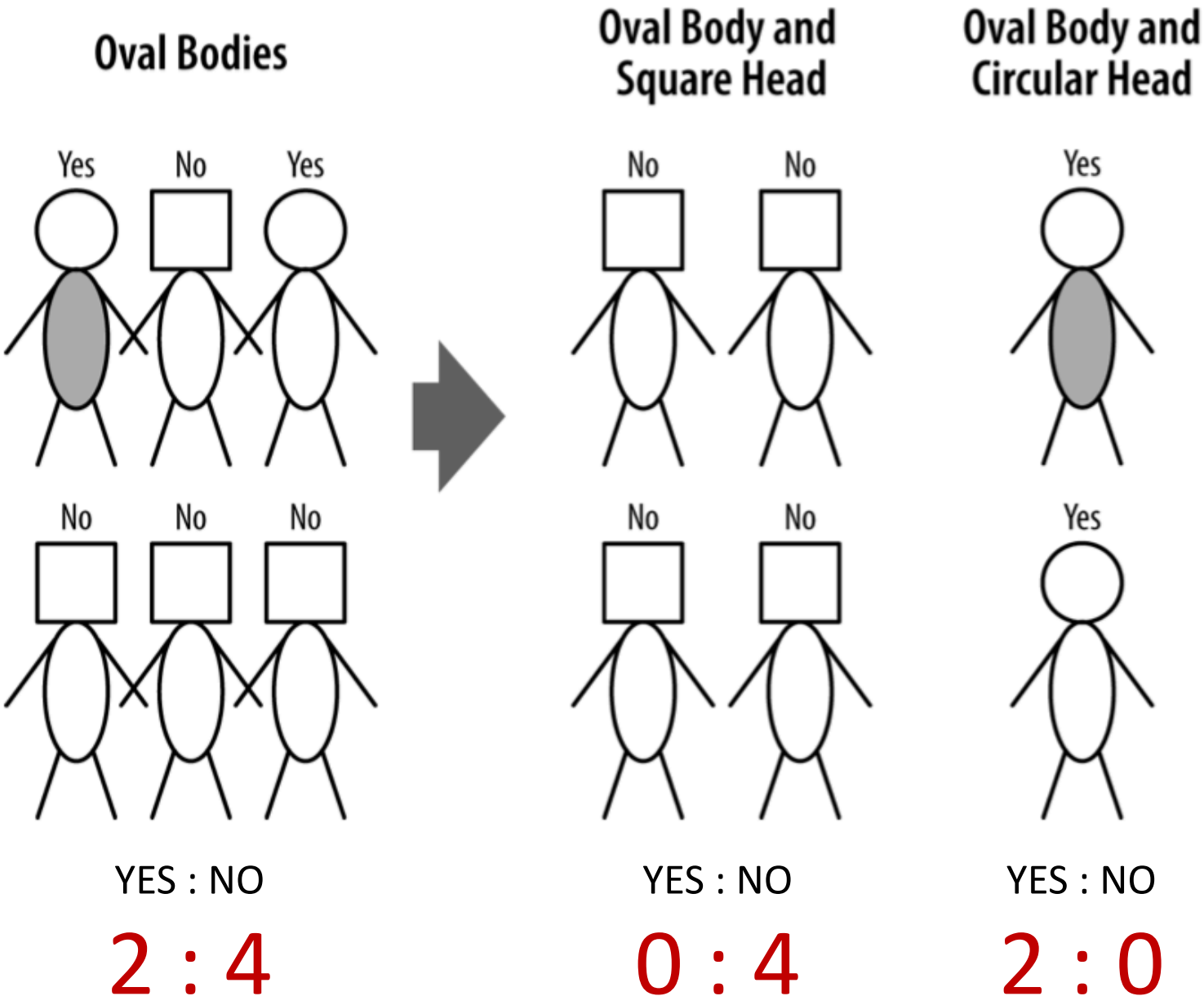
5 : 1

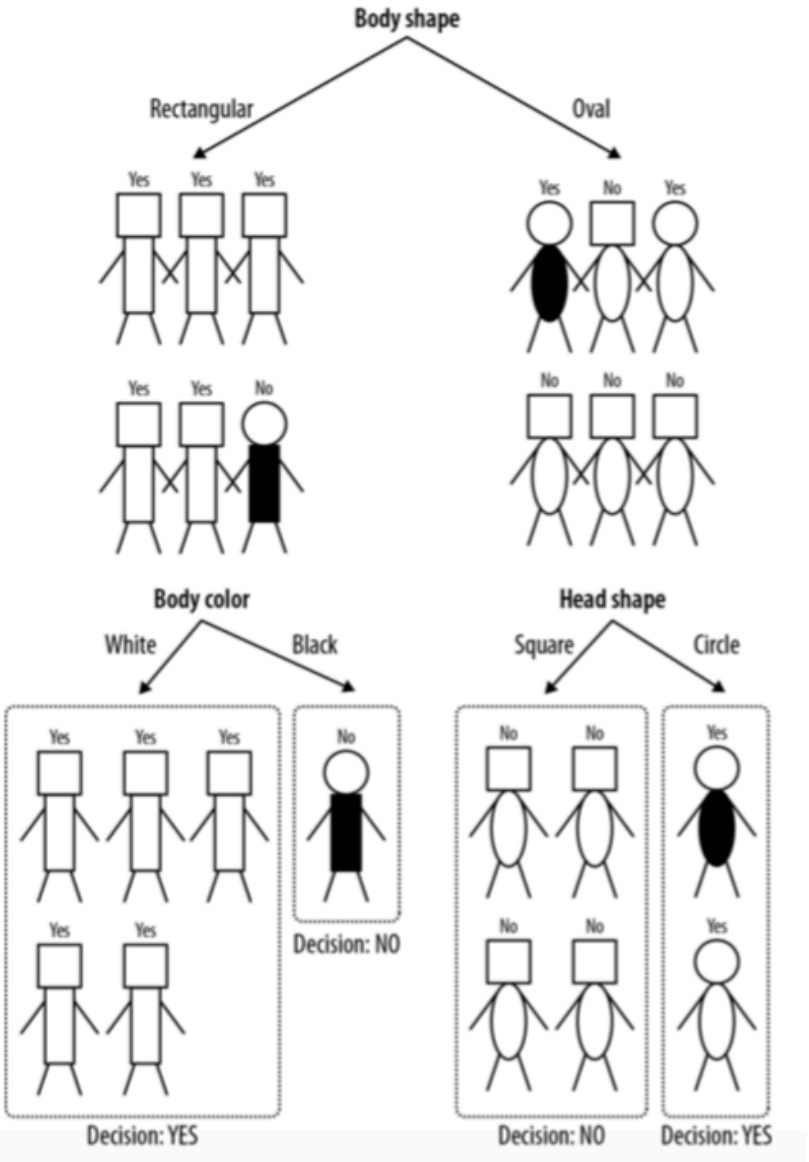


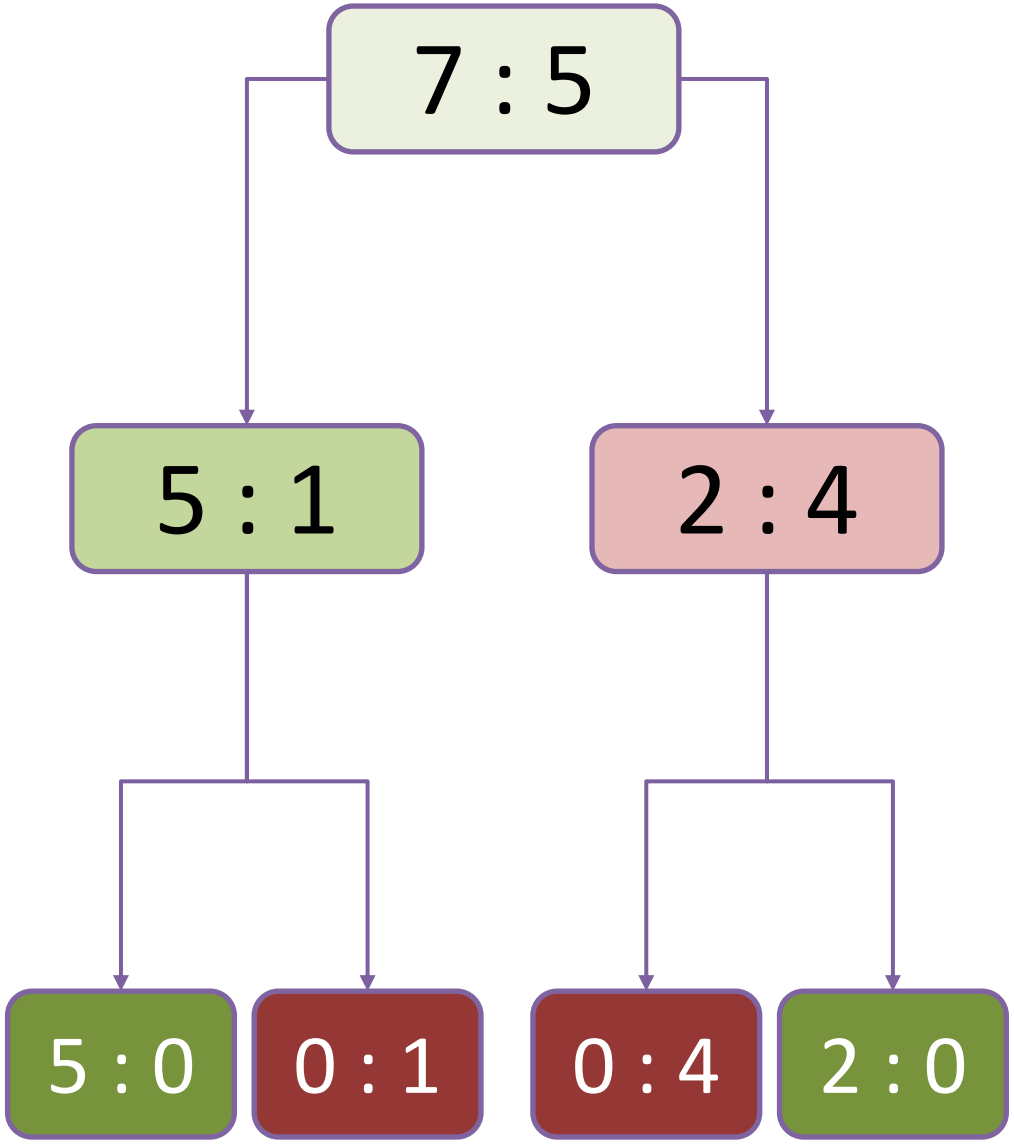
YES : NO

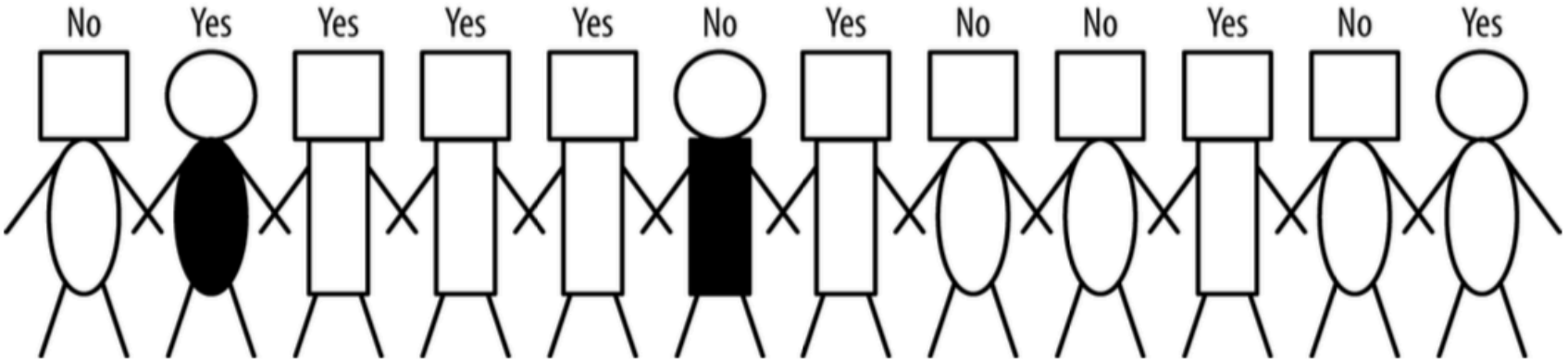
2 : 4


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

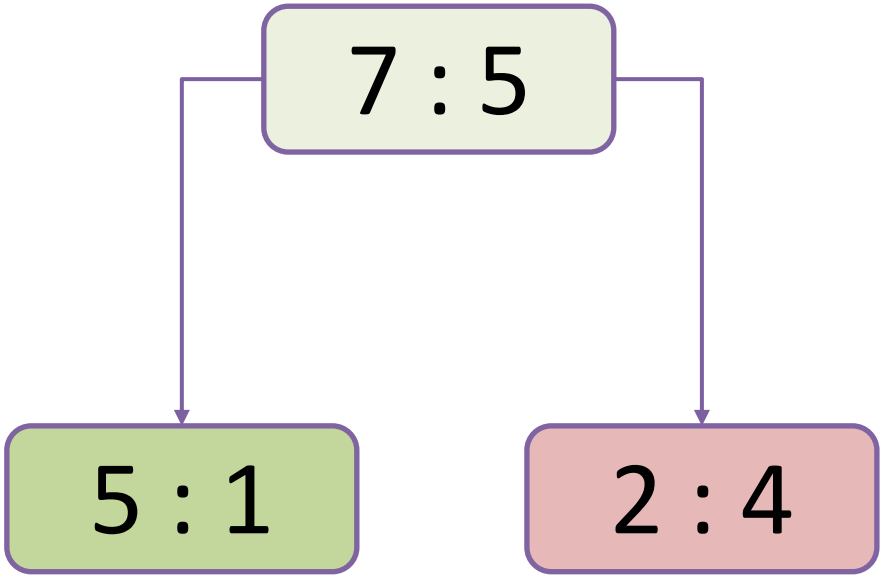

Oval Bodies
Oval Body and Square Head
Oval Body and Circular Head
Yes
No
Yes
No
No
No
No
No
No
No
No
Yes
YES : NO
2 : 4
YES : NO
0 : 4
YES : NO
2 : 0


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

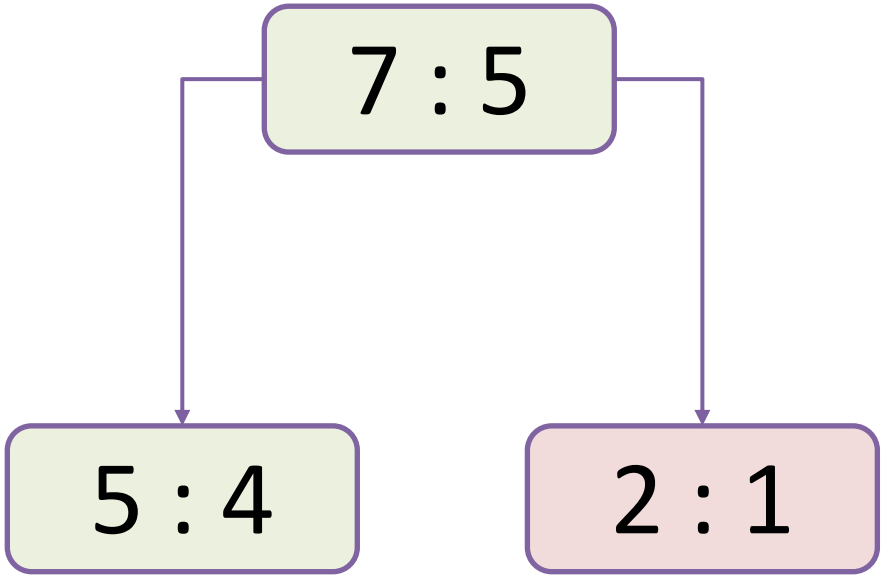

7 : 5
5 : 1
2 : 4
5 : 0
0 : 1
0 : 4
2 : 0



Body : Rectangle or Oval



Head : Square or Circle

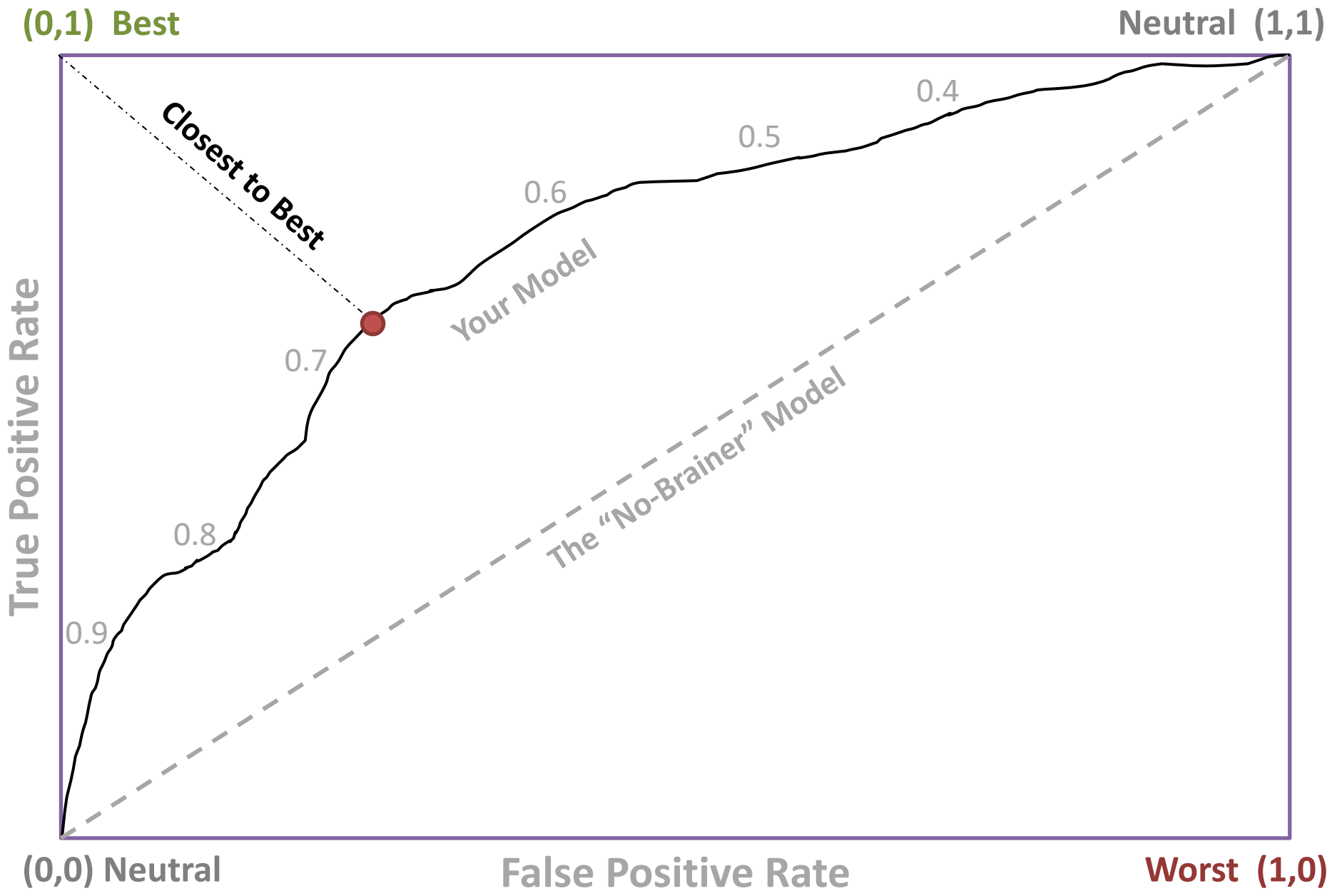


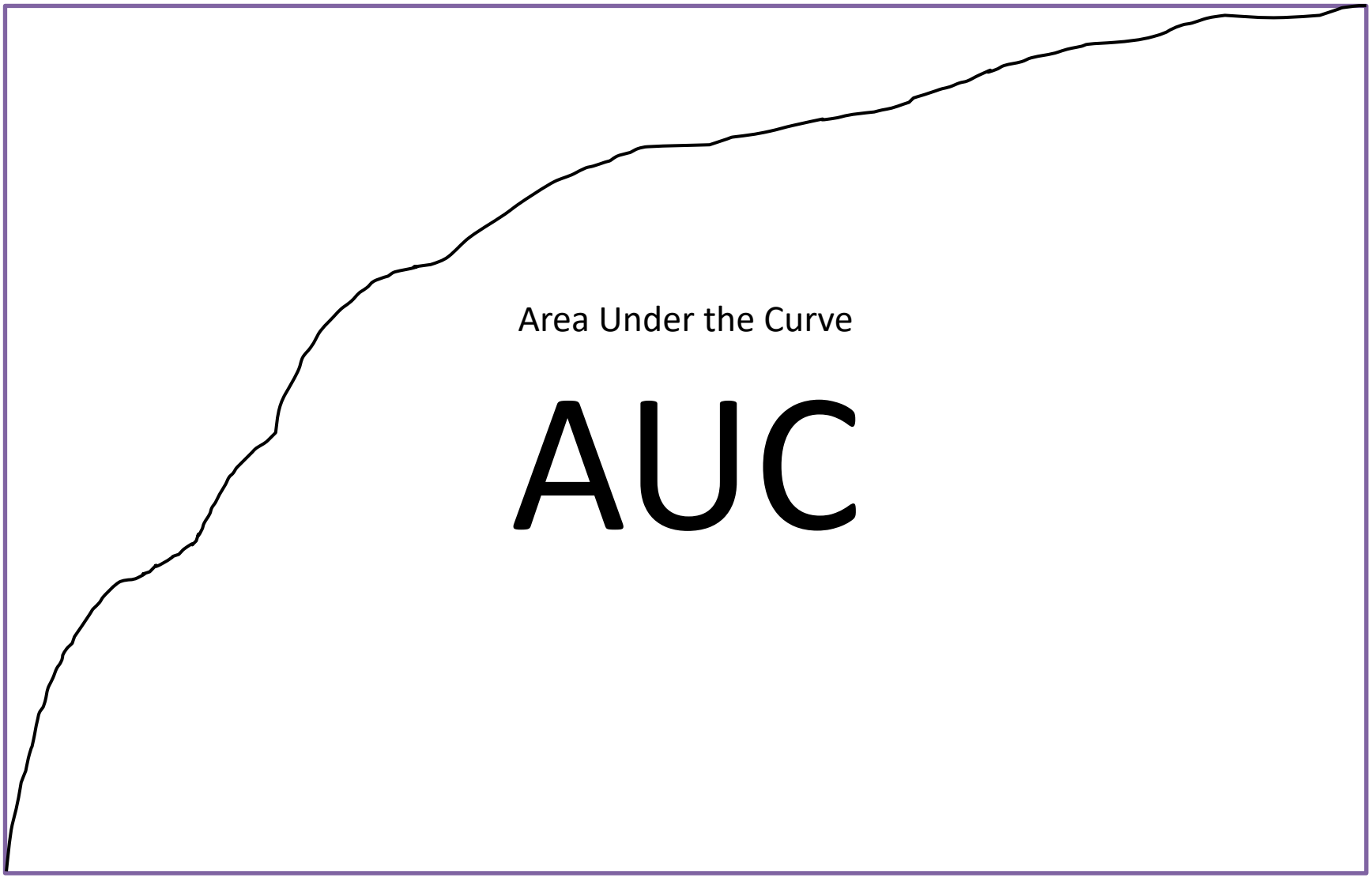
Prediction using a Decision Tree

HOW ACCURATE IS IT?

Confusion Matrix

	Predicted **Bad**	Predicted **Good**
Actually **Bad**	True Negatives	False Positives
Actually **Good**	False Negatives	True Positives


(0,1) Best
Neutral (1,1)
Closest to Best
0.4
0.5
0.6
Your Model
0.7
The "No-Brainer" Model
0.8
0.9
True Positive Rate
(0,0) Neutral
False Positive Rate
Worst (1,0)


Area Under the Curve
AUC