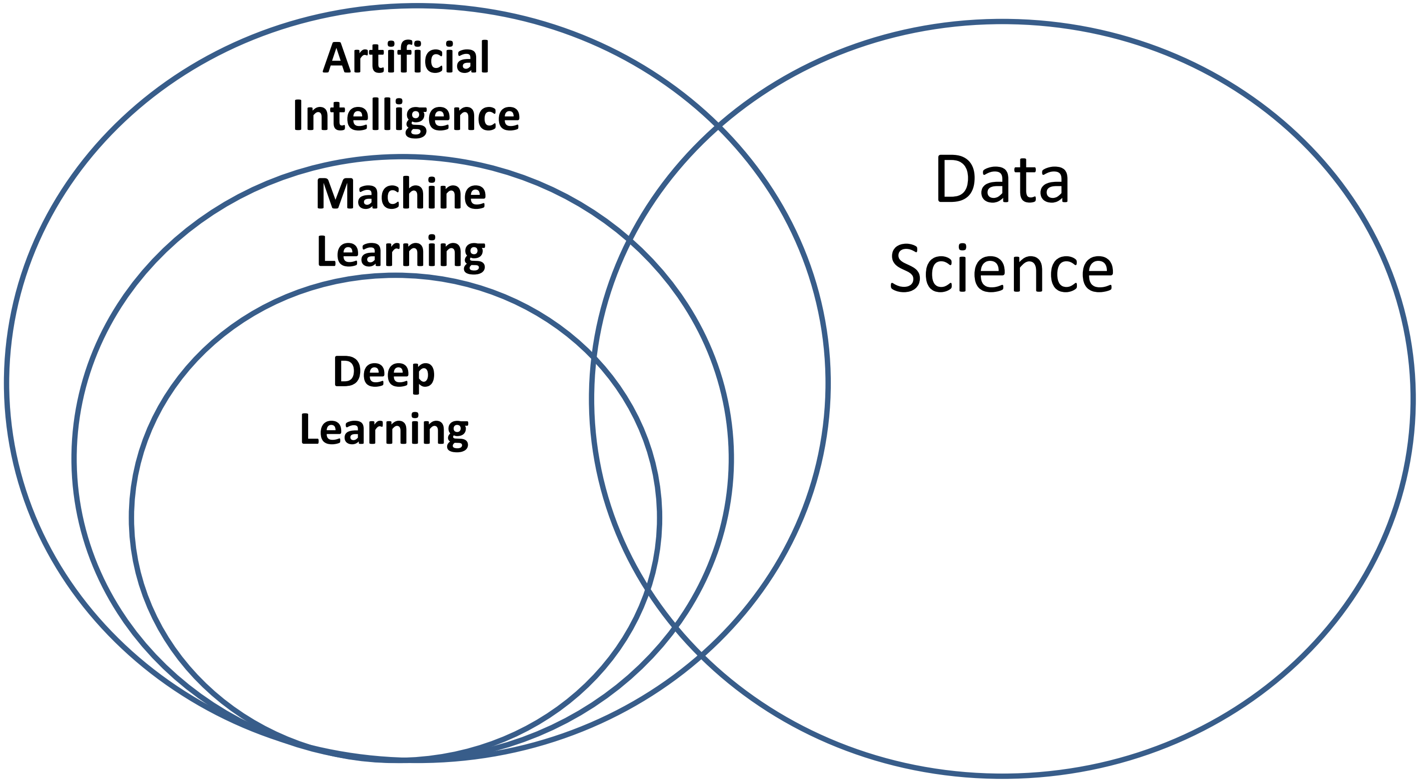


Artificial Intelligence : Human Intelligence exhibited by machine.

Narrow AI : Computers can do a specific/one thing very well.

General AI : Computers can do multiple things like humans. We are very far away from this.

Machine Learning : Approach to try and achieve AI through systems that can find patterns in data. Stanford Univ - Science of getting computers to act without being explicitly programmed.

Deep Learning : One of the techniques to implement machine learning.

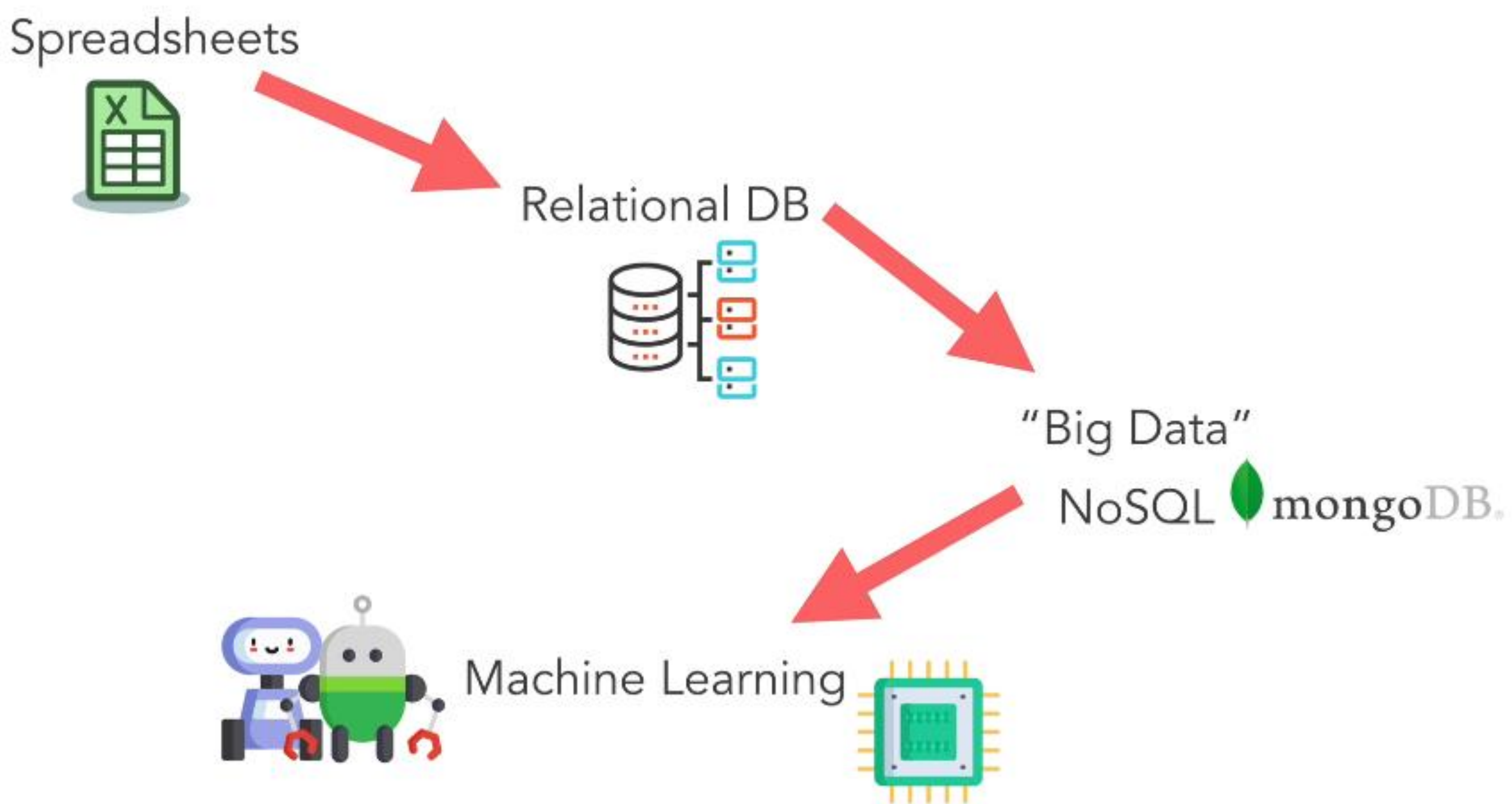
Data Science : Analysing Data

Play Ground

- https://teachablemachine.withgoogle.com/

- https://ml-playground.com/#

How did we get here ?



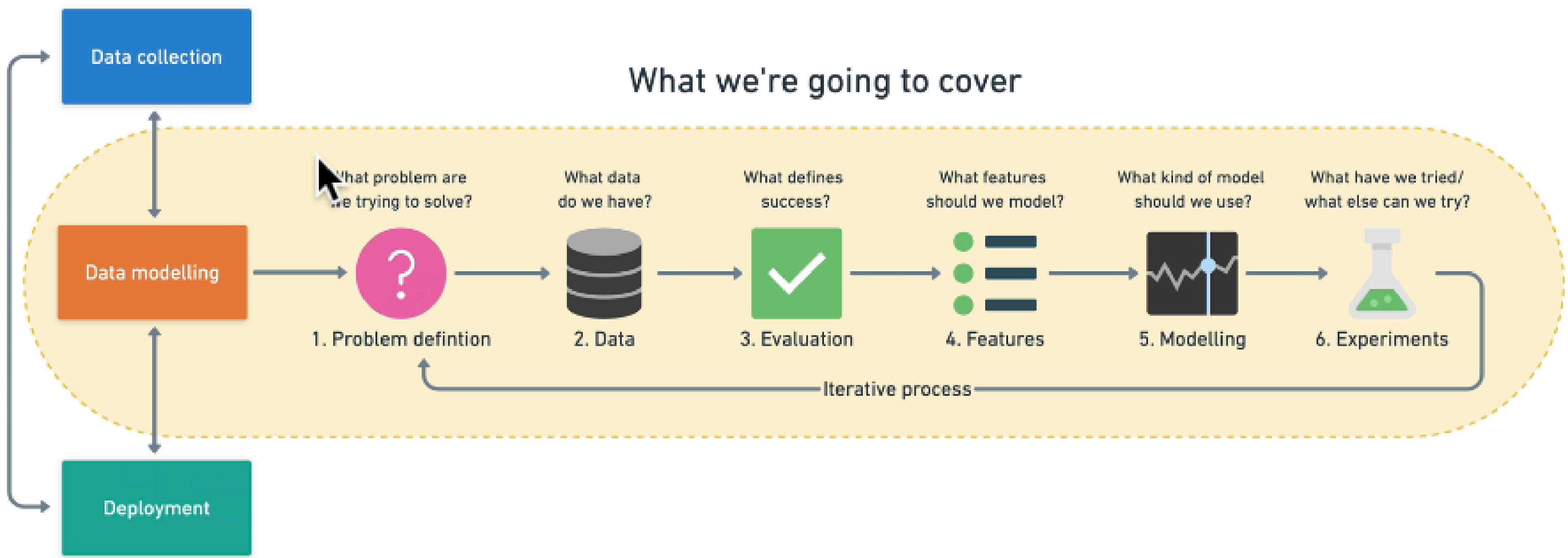
YouTube Recommendation Engine

- https://ml-playground.com/#

- X Axis - Duration of Video
- Y Axis - Likes to the Video

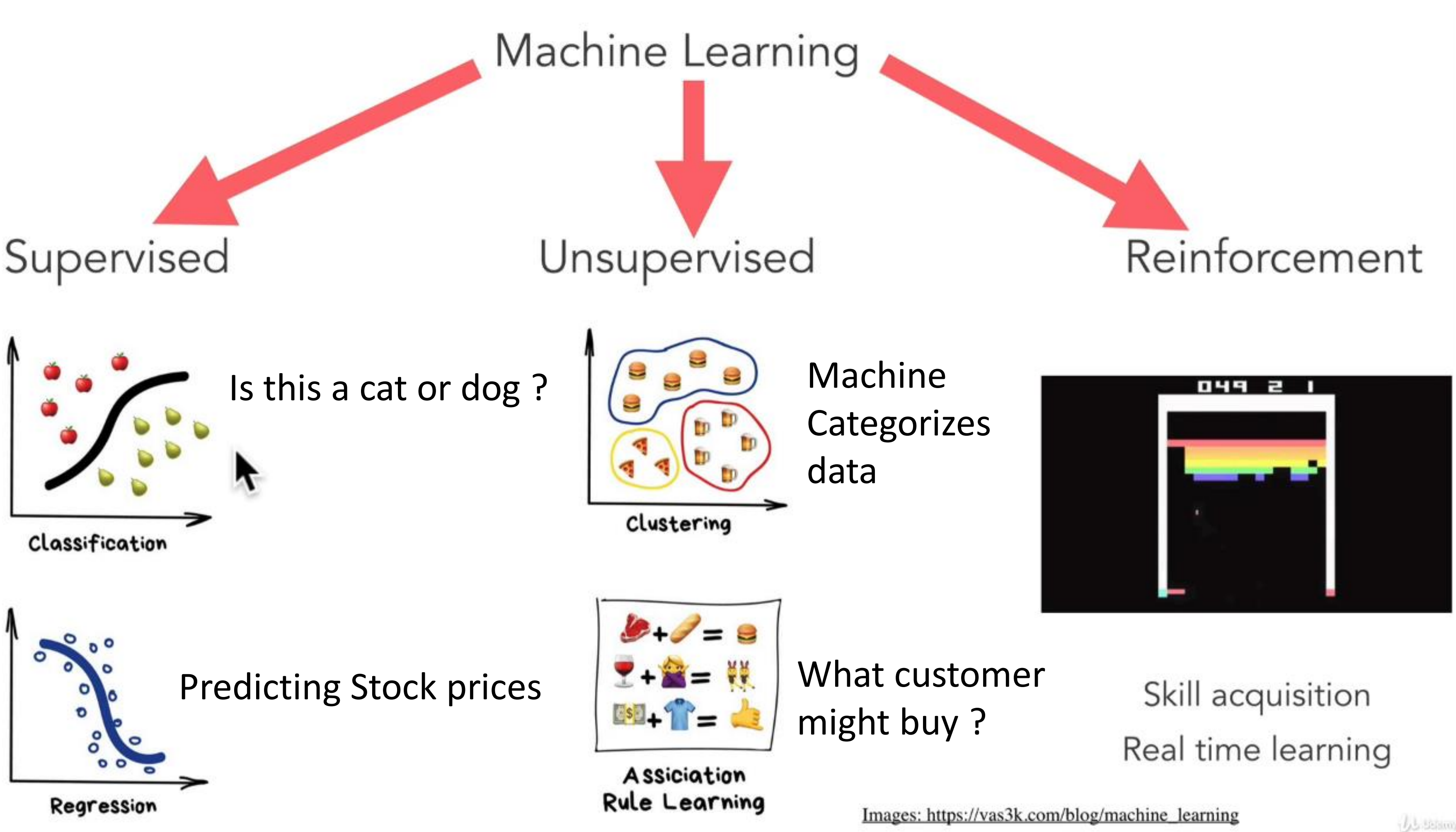
Framework

Steps in a full machine learning project



1. Problem Definition

Types of machine learning



When not to use machine learning ?

- Will simple hand coded instructions based system work ? If yes, then use it.

Main types of machine learning

Supervised learning

Classification

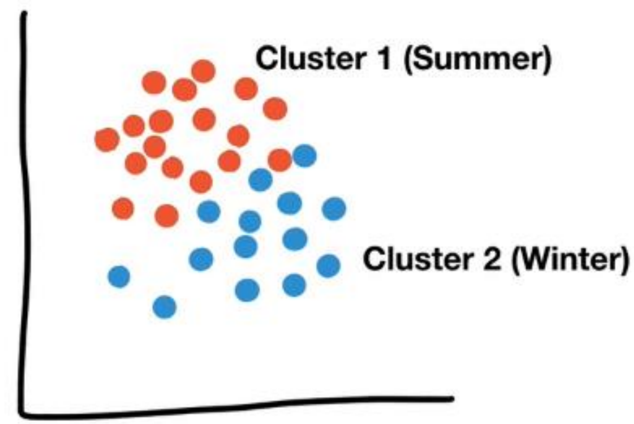
- "Is this example one thing or another?"
- Binary classification = two options
- Multi-class classification = more than two options

Regression

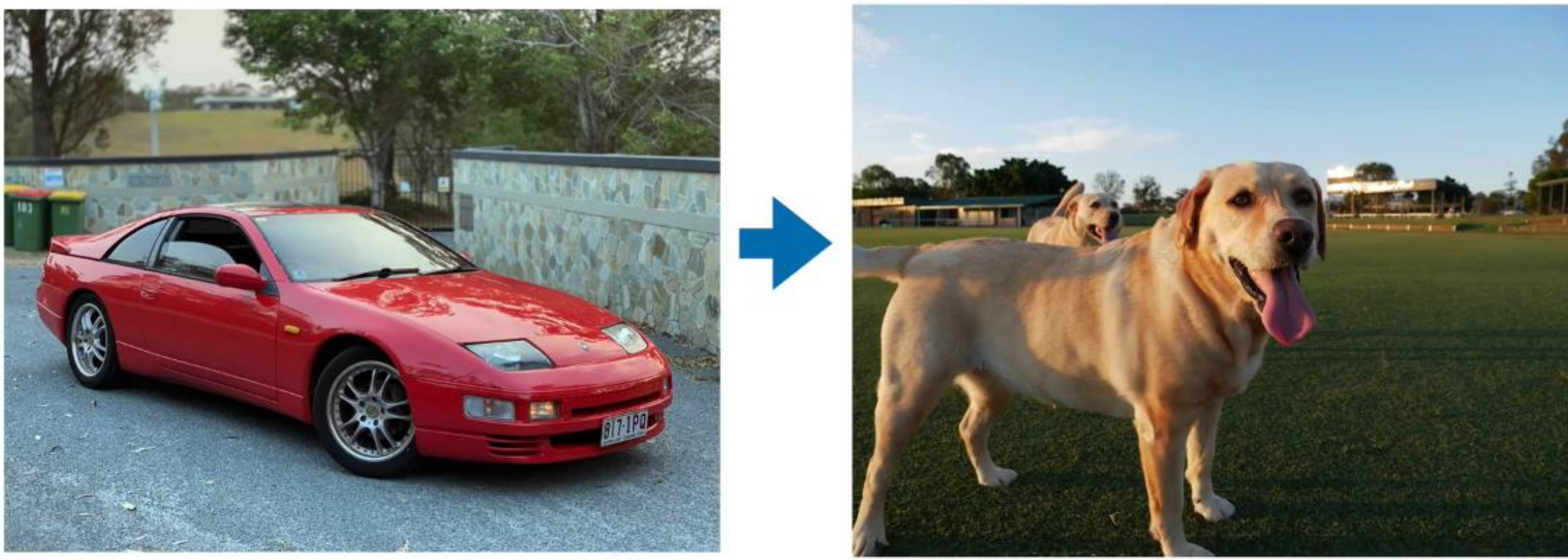
- "How much will this house sell for?"
- "How many people will buy this app?"

Unsupervised learning

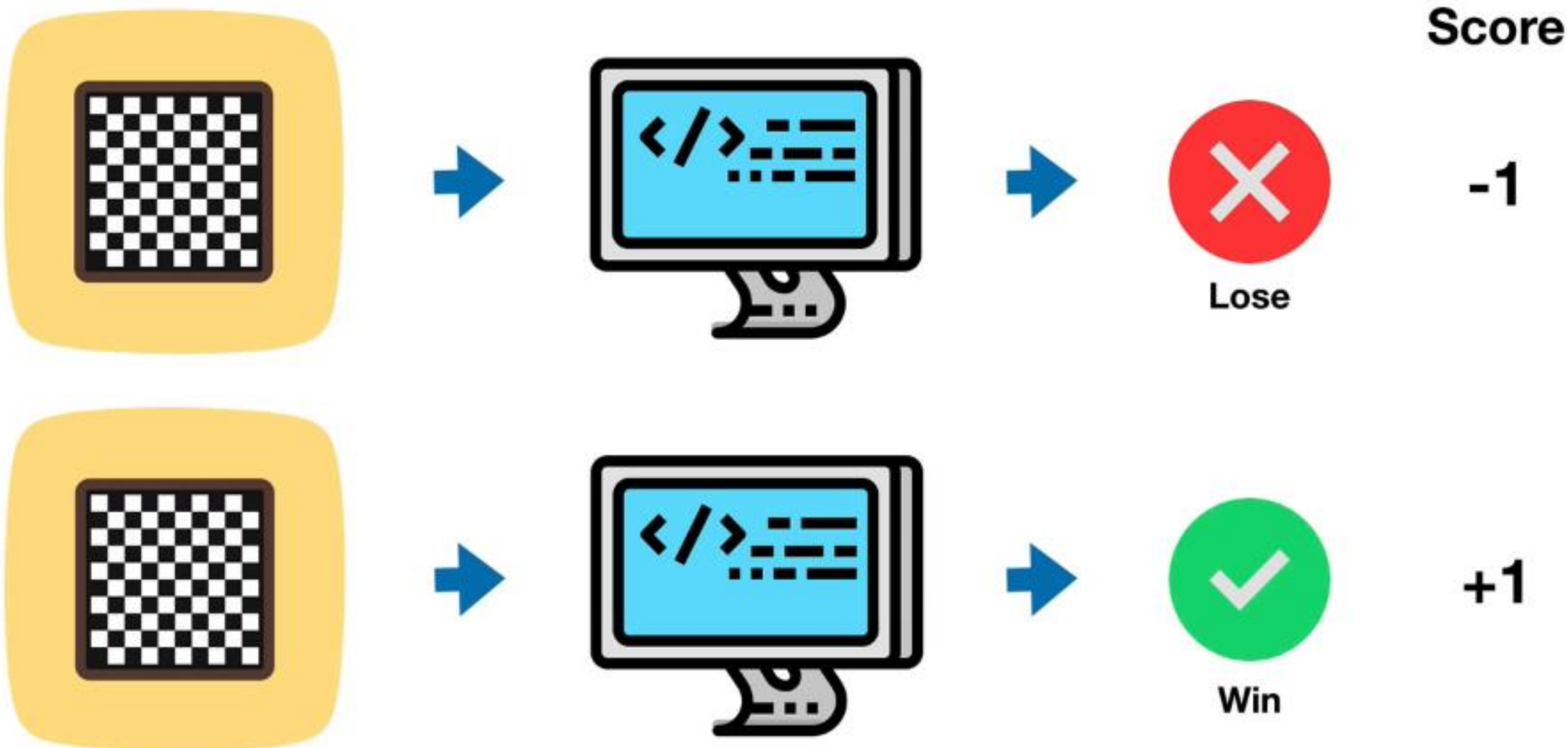
Customer ID	Purchase 1	Purchase2
1	Sunglasses	Singlet
2	Jacket	Snow boots
3	Sunscreen	Beach towel



Transfer Learning



Reinforcement Learning

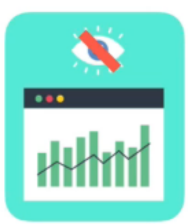


Problem Definition

Matching your problem

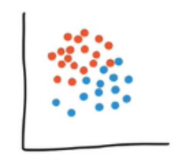
Supervised Learning

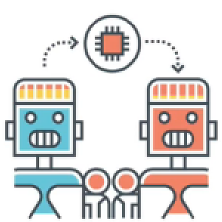
"I know my inputs and outputs."



Unsupervised Learning

"I'm not sure of the outputs but I have inputs."





Transfer Learning

"I think my problem may be similar to something else."

2.Data

“What kind of data do we have?”

Types of Data

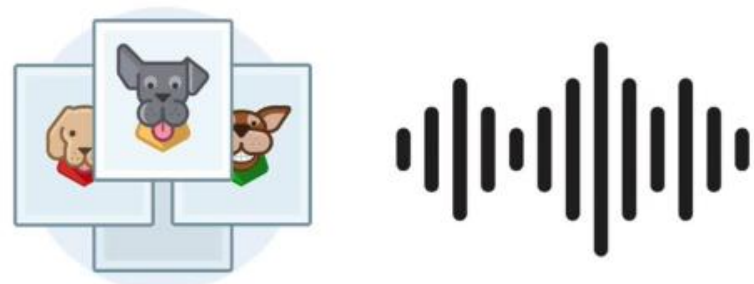
Rows

Columns

ID	Weight	Sex	Blood Pressure	Chest pain	Heart disease?
4328	110Kg	M	120/80	4	Yes
5681	64Kg	F	130/90	1	NO
7911	81Kg	M	130/80	0	NO

Table 1.0: Patient records

Structured

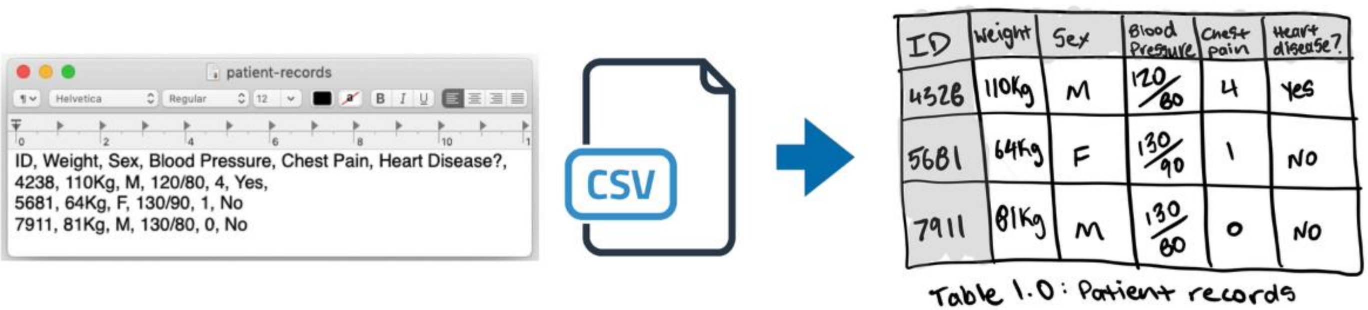


From: daniel@mrdbourke.com
Hey Daniel,

First of all, thank you for being so amazing.
This machine learning course is incredible.
Thank you for keeping it simple!

Unstructured

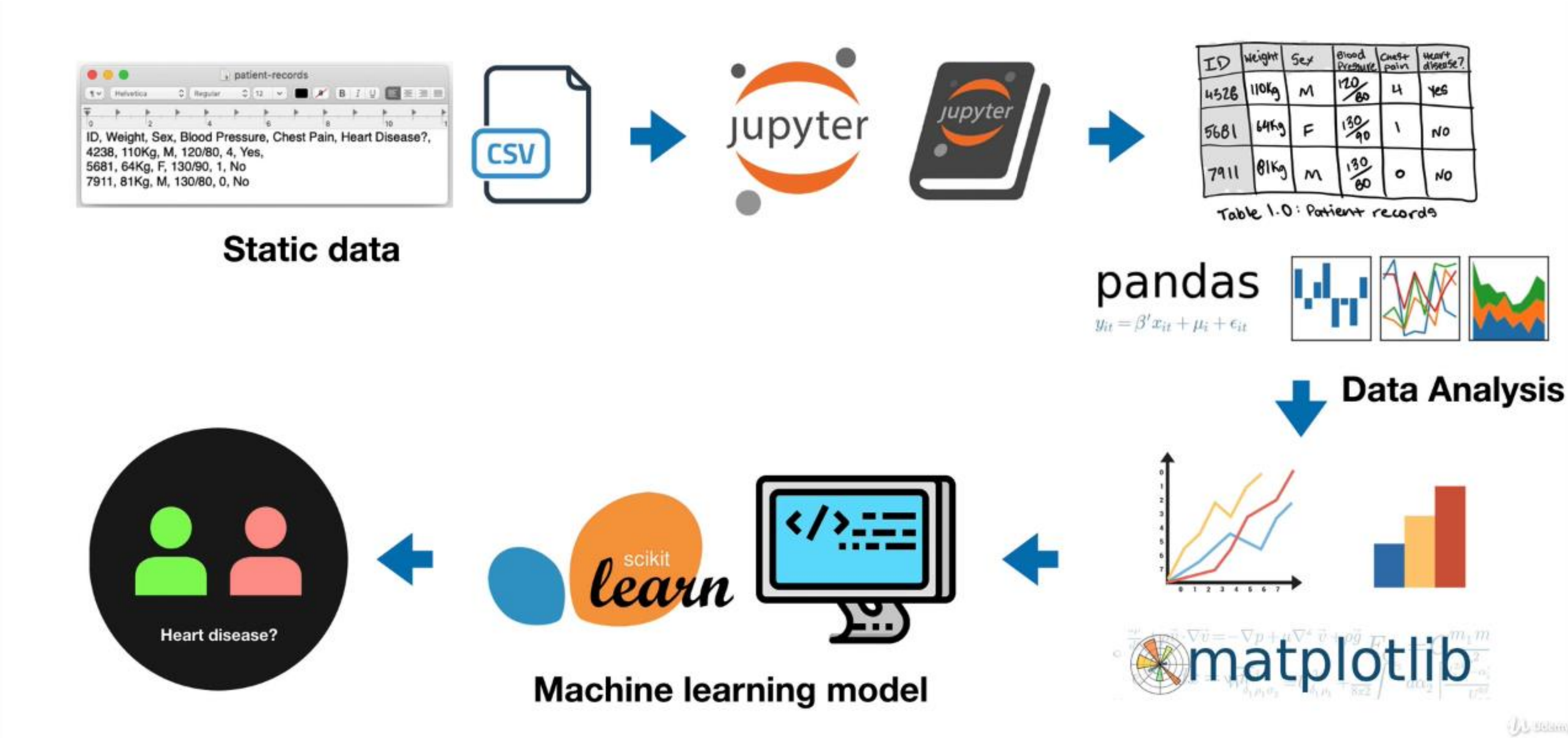
Types of Data ....



Static

Streaming

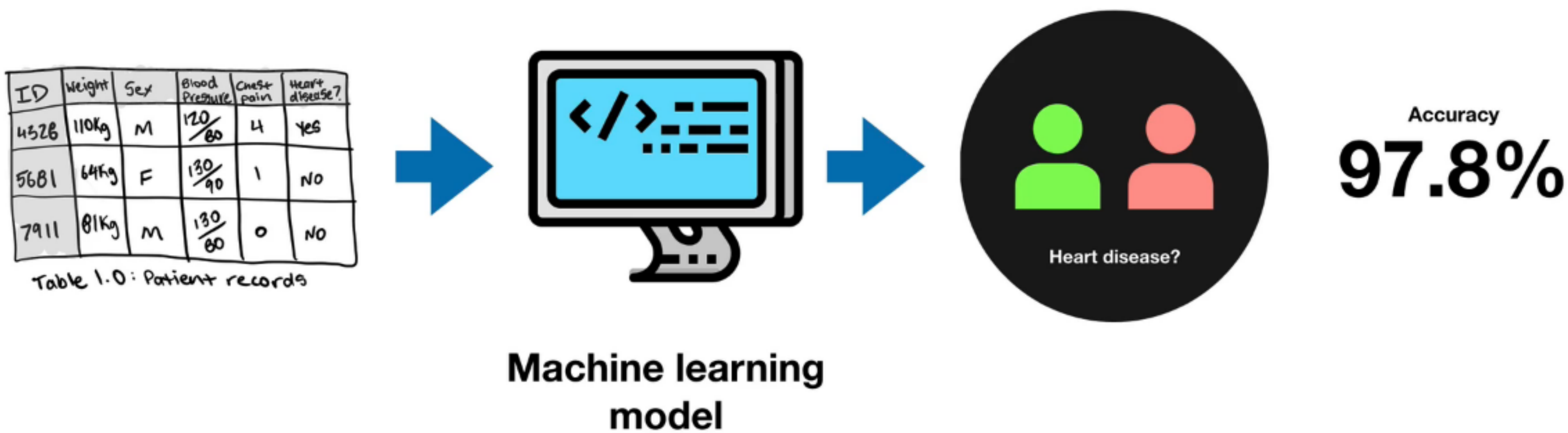
A data science workflow



3. Evaluation ✅

"What defines success for us?"

“For this project to be worth pursuing further, we need a machine learning model with over 99% accuracy.”



Types of metrics

Classification	Regression	Recommendation
Accuracy	Mean absolute error (MAE)	Precision at K
Precision	Mean squared error (MSE)	
Recall	Root mean squared error (RMSE)	

Classifying Car insurance claims

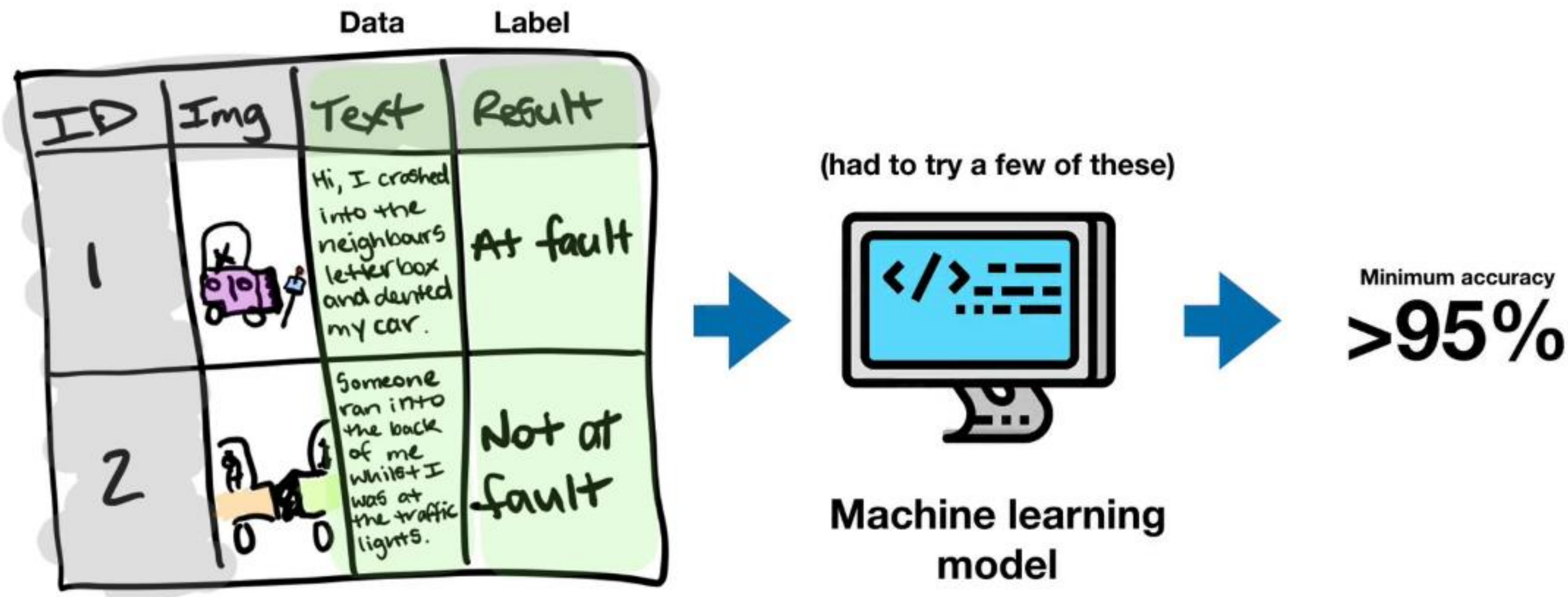


Table 2.0: Car insurance claims

4. Features

"What do we already know about the data?"

Feature variables can be

- Numerical
- Categorical

Feature engineering

- Deriving new features from existing one.

Feature Coverage

- Checking if values are correctly populated for a feature or not ? Do not use it if it is not well covered.

	Feature variables				Target variable	Derived feature	
ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?	visit in last year?	Most eaten food
4326	110Kg	M	81	4	Yes	Yes	Fries
5681	64Kg	F	61	1	NO	Yes	?
7911	81Kg	M	57	0	NO	NO	?

Table 1.0: Patient records

5. Modelling Part 1 — 3 sets

"Based on our problem and data, what model should we use?"

3 parts to modelling

1. Choosing and training a model

or

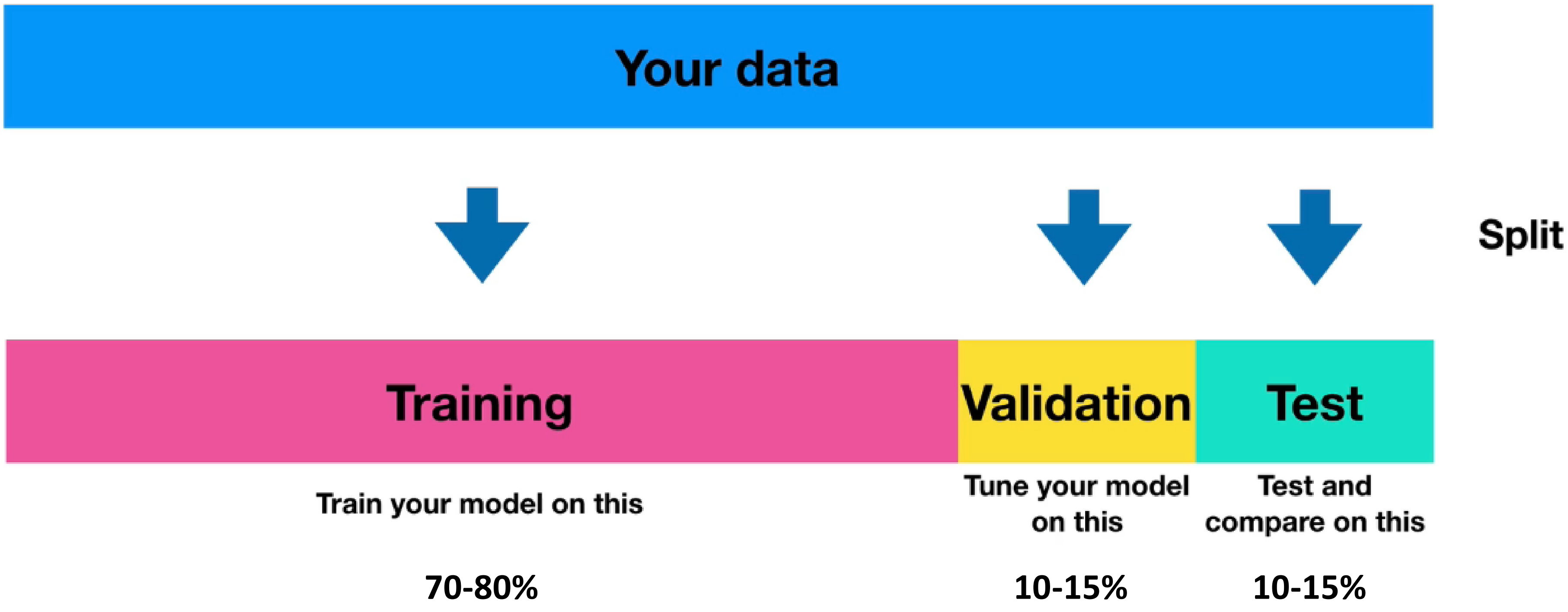
2. Tuning a model

3. Model comparison

vs.

vs.

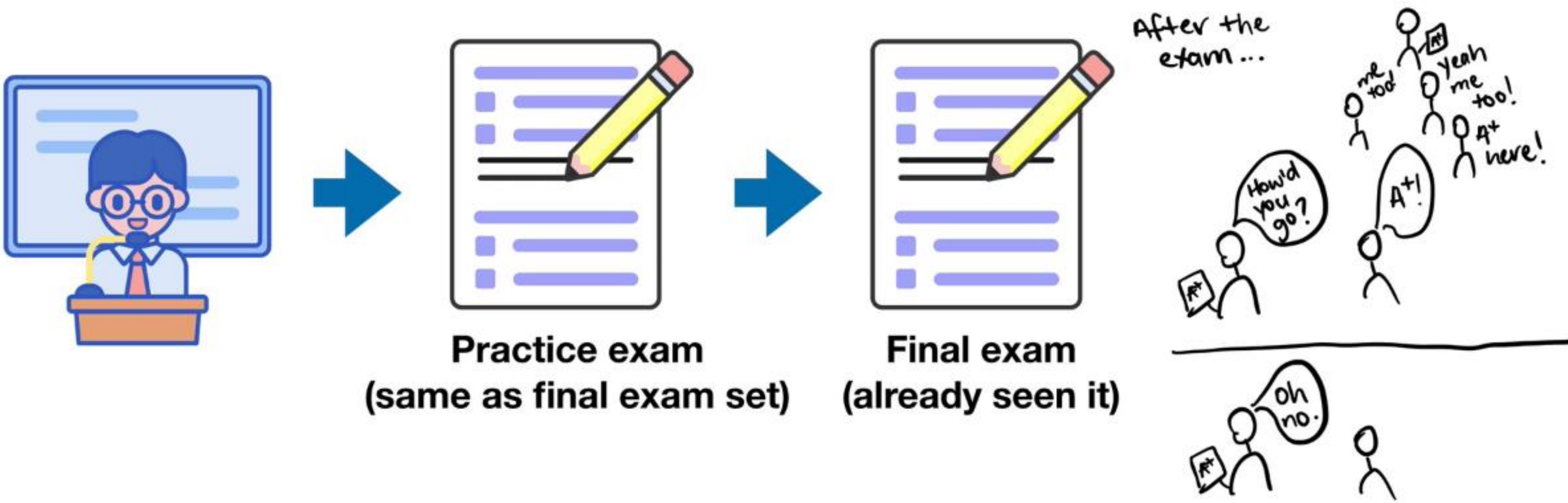
Training, validation and test sets



3 sets

Generalization – The ability for a machine learning model to perform well on data it hasn't seen before.

When things go wrong ?

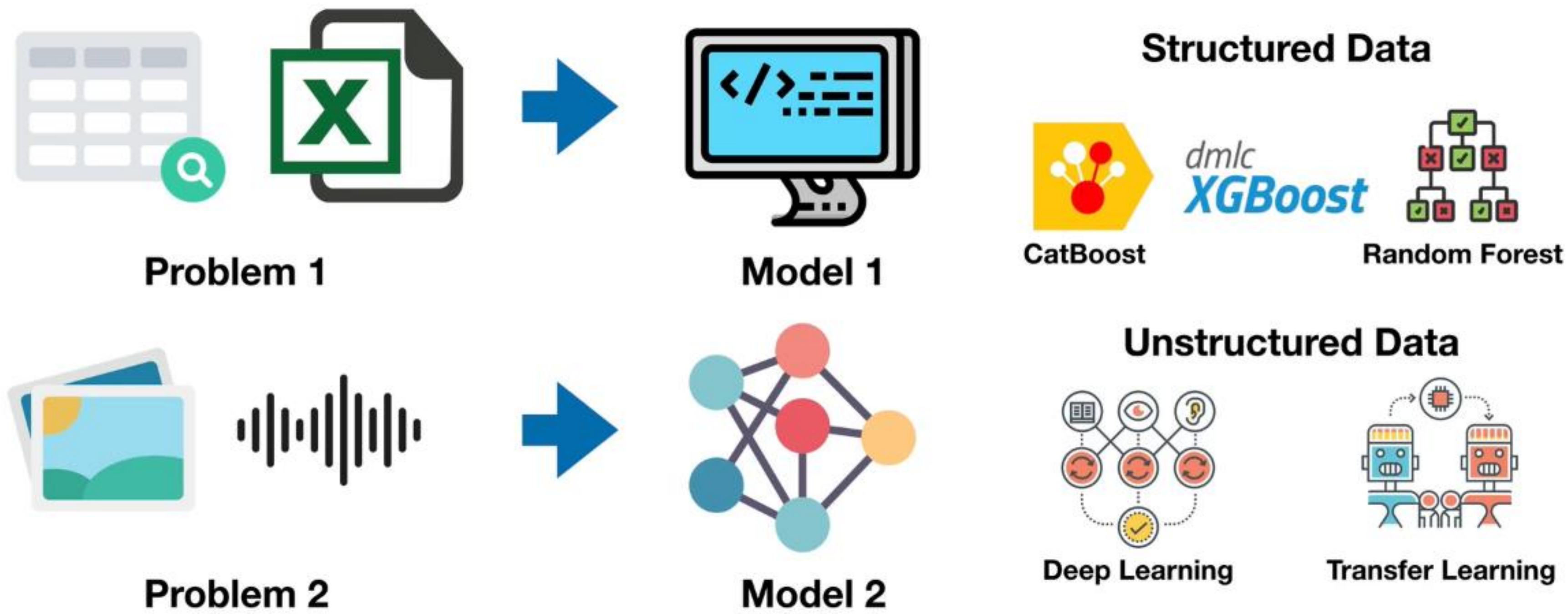


Machine really did not learn anything, it just memorized what it solved in training part.

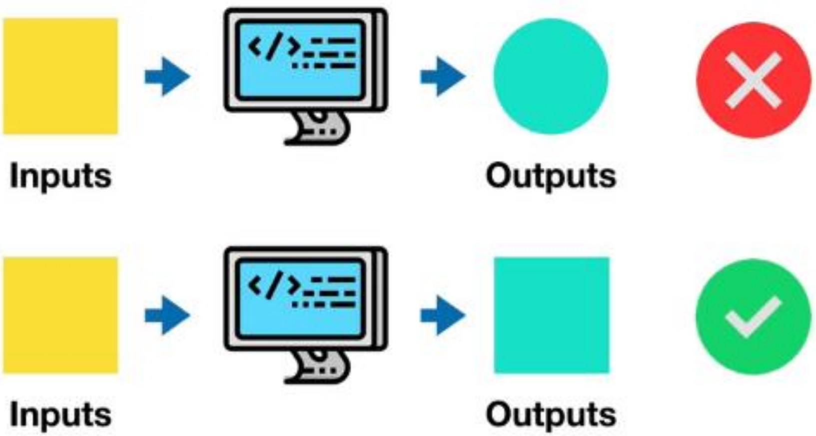
5. Modelling Part 2 — Choosing

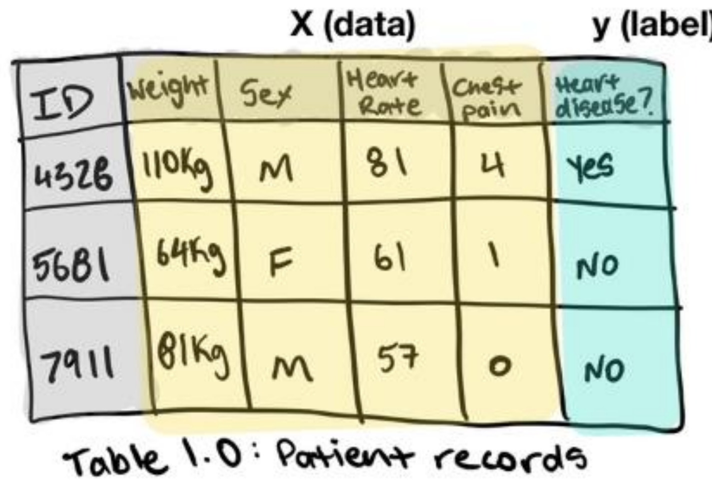
“Based on our problem and data, what model should we use?”

Choosing a model



Training a model





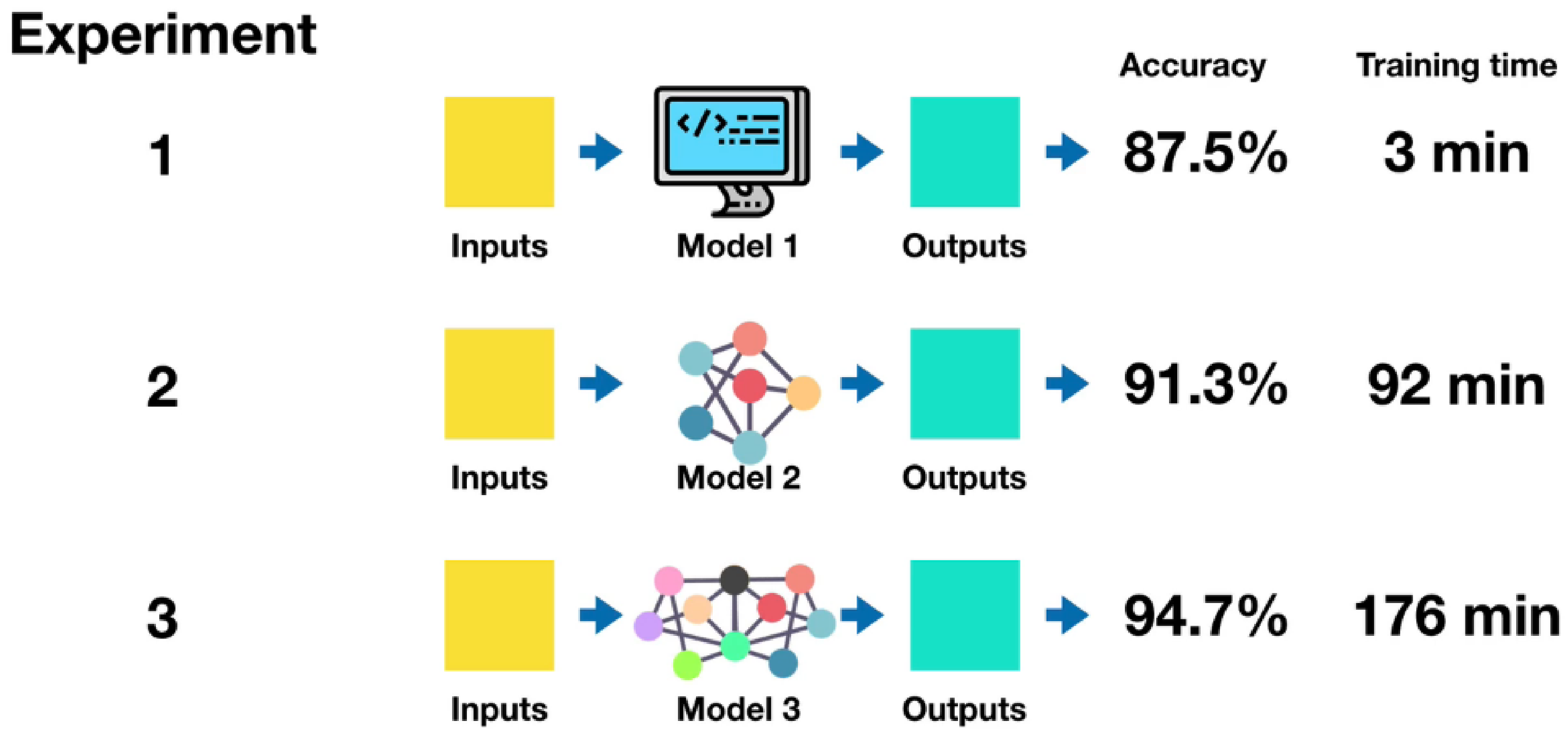
X (data) | y (label)

ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?
4328	110Kg	M	81	4	Yes
5681	64Kg	F	61	1	NO
7911	81Kg	M	57	0	NO

Table 1.0: Patient records

Training Data

Goal - Minimize time between experiments



Sometimes for smaller %age extra of Accuracy, we end up spending lot of time. We should avoid that.

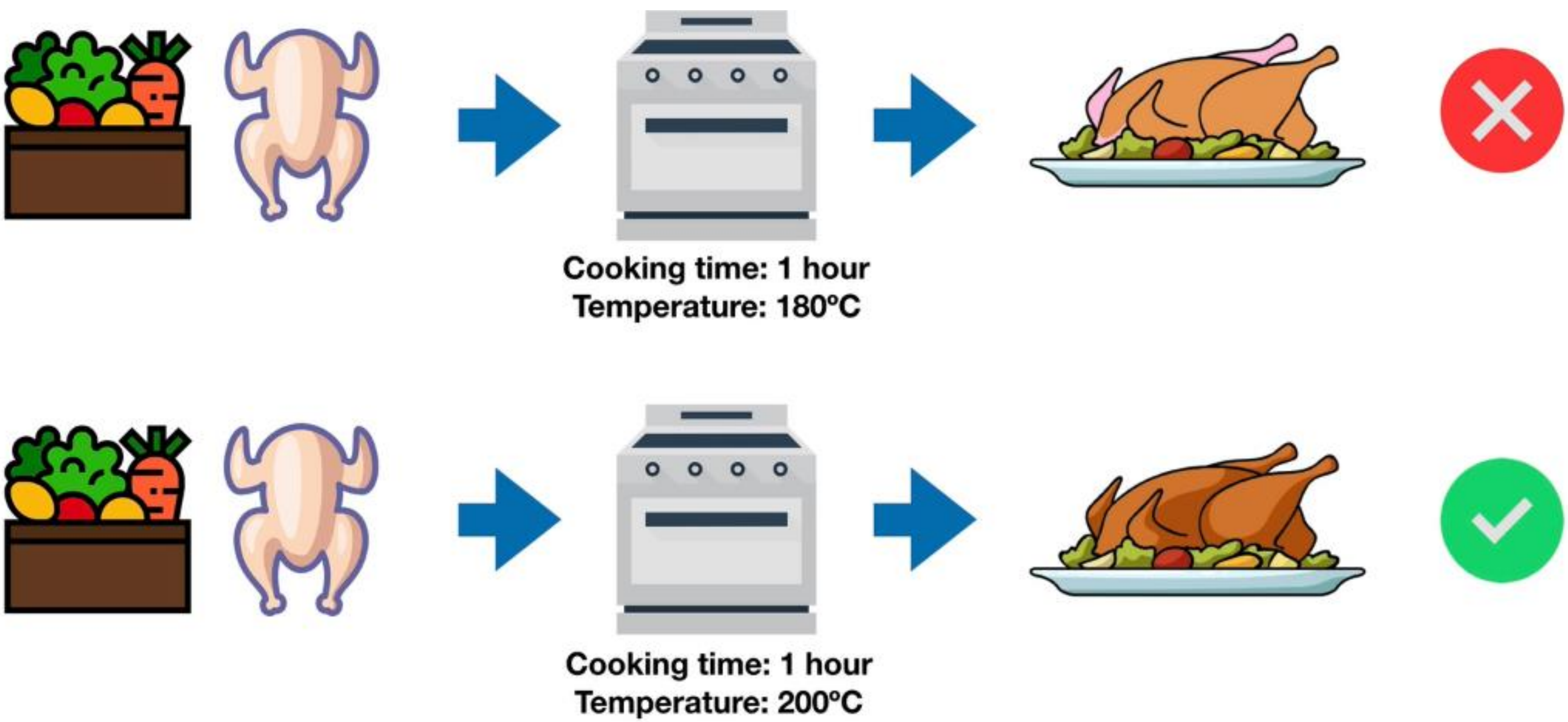
Remember

- Some models work better than others on different problems
- Don't be afraid to try things
- Start small and build up (add complexity) as you need

5. Modelling Part 3 — Tuning

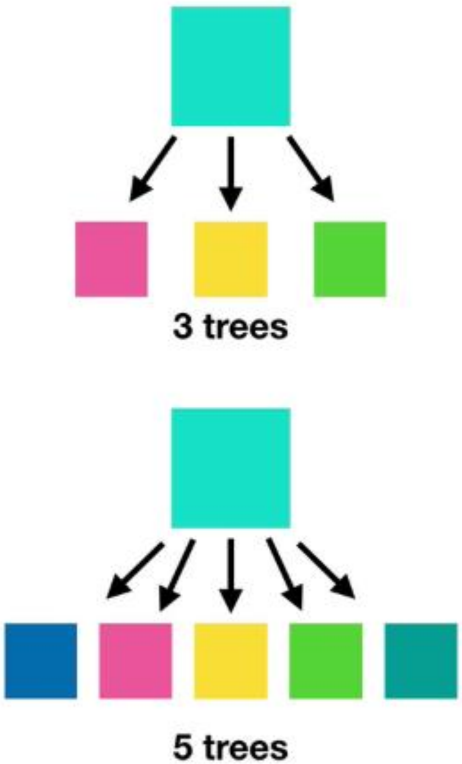
“Based on our problem and data, what model should we use?”

Tuning

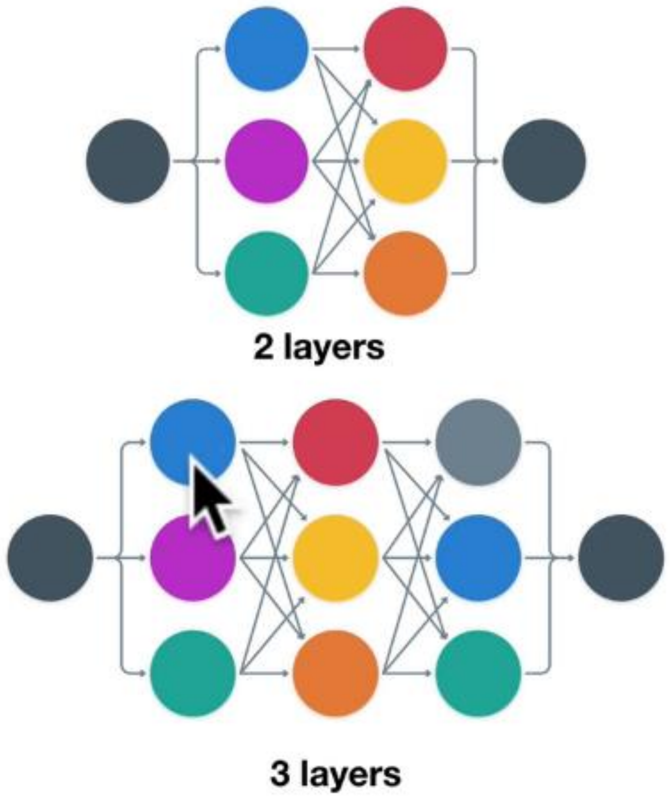

Cooking time: 1 hour
Temperature: 180ºC
Cooking time: 1 hour
Temperature: 200ºC

Tuning…

Random Forest



Neural Networks



Remember

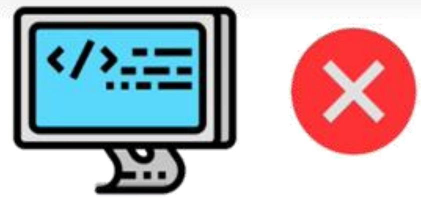
- Machine learning models have hyperparameters you can adjust
- A models first results aren’t its last
- Tuning can take place on training or validation data sets

5. Modelling Part 4 — Comparison

“How will our model perform in the real world?”

Model performance

Data Set	Performance
Training	98%
Test	96%



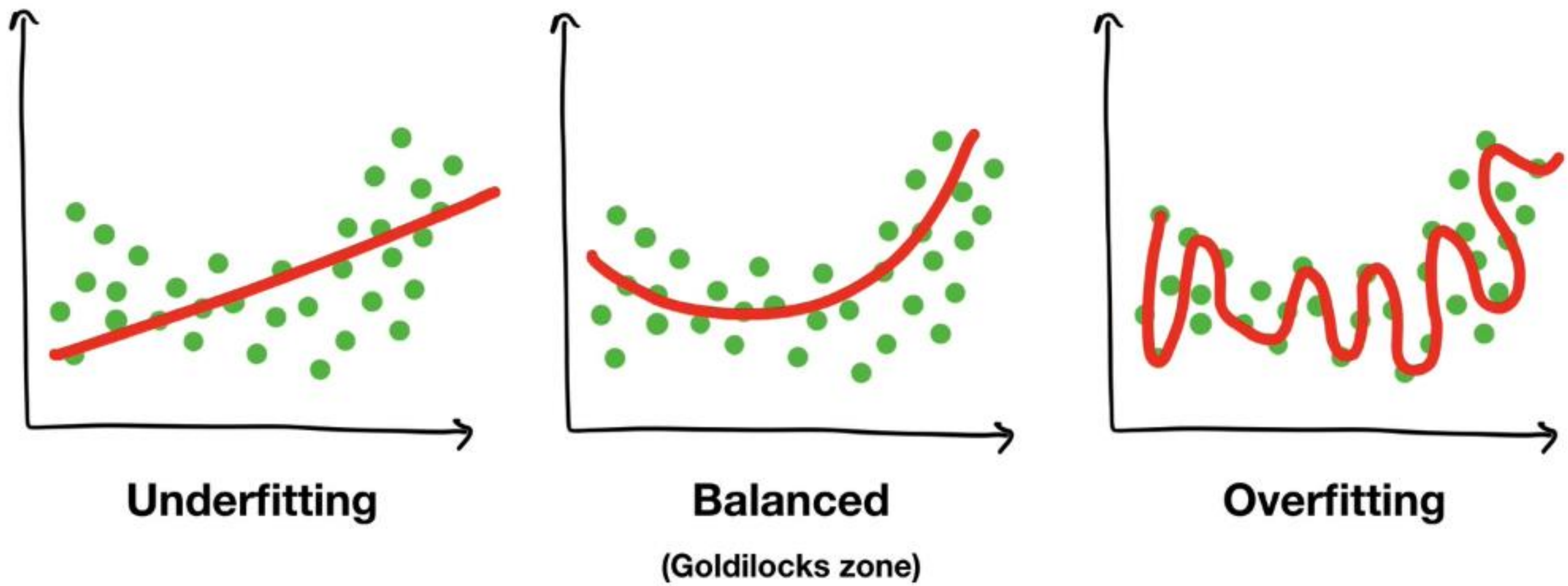
Underfitting (potential)

Data Set	Performance
Training	64%
Test	47%

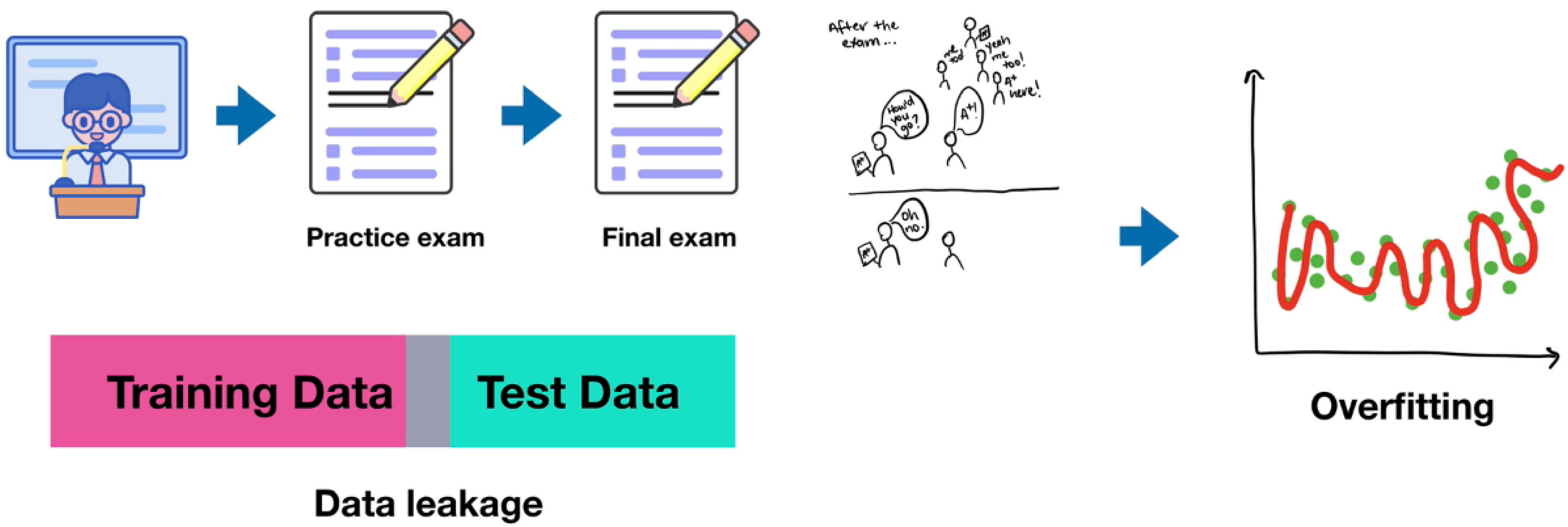
Overfitting (potential)

Data Set	Performance
Training	93%
Test	99%

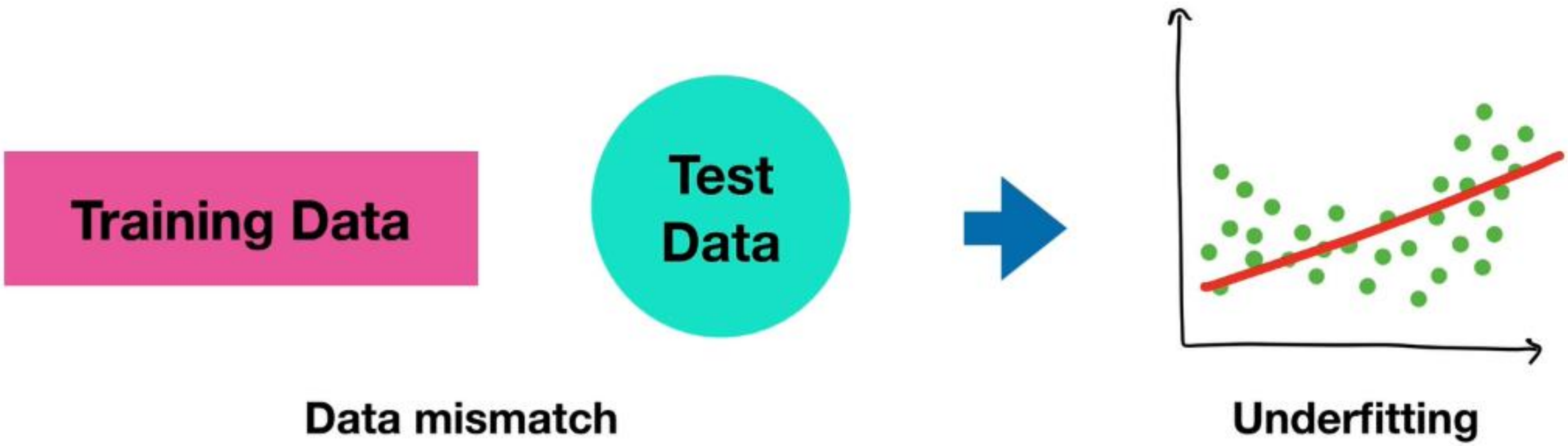
Overfitting and Underfitting



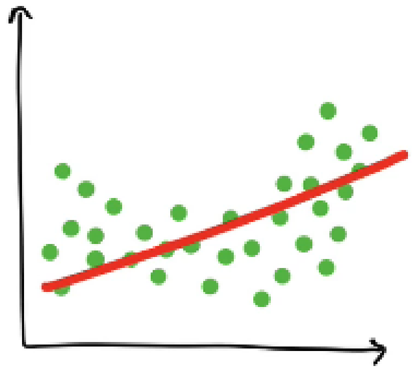
Overfitting and Underfitting



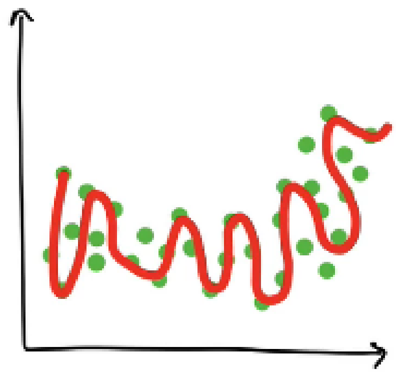
Overfitting and Underfitting



Fixes for Overfitting and Underfitting



Underfitting

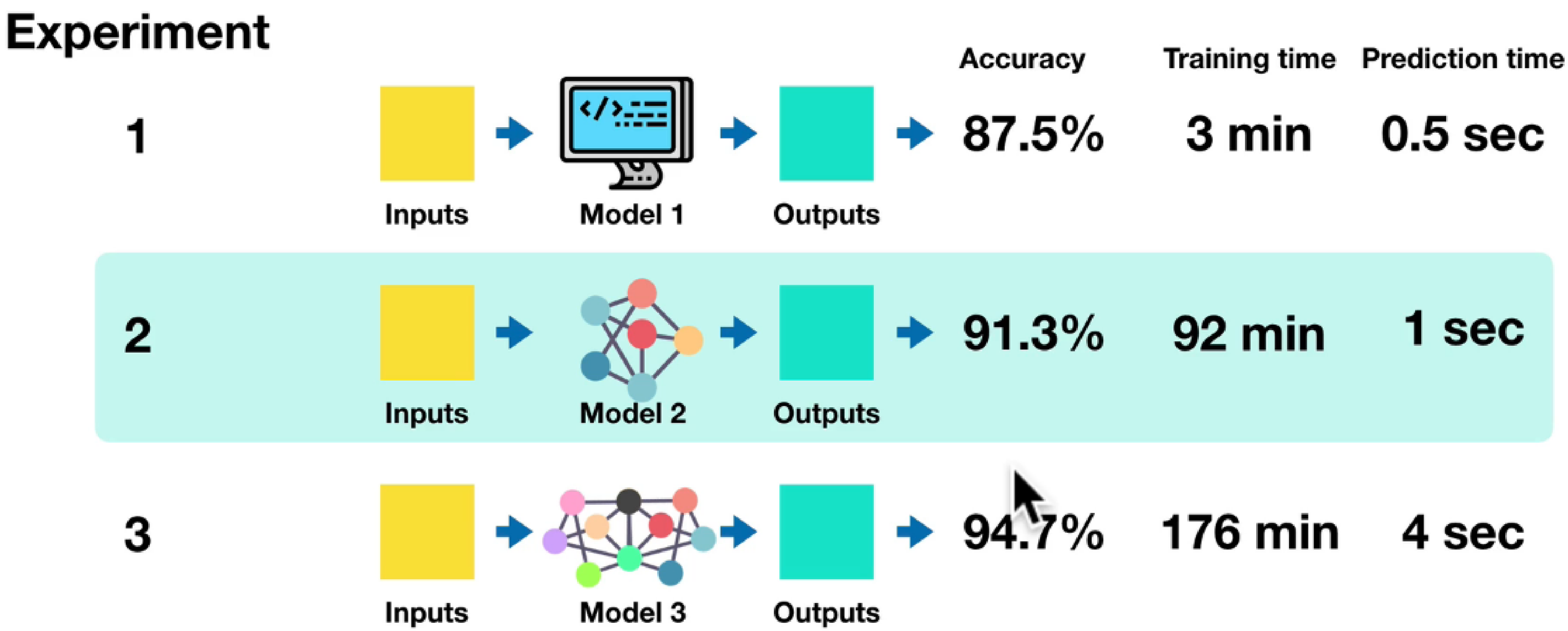


Overfitting

- Try a more advanced model
- Increase model hyperparameters
- Reduce amount of features
- Train longer

- Collect more data
- Try a less advanced model

Comparison


Experiment
Accuracy
Training time
Prediction time
1
Inputs
Model 1
Outputs
87.5%
3 min
0.5 sec
2
Inputs
Model 2
Outputs
91.3%
92 min
1 sec
3
Inputs
Model 3
Outputs
94.7%
176 min
4 sec

Remember

- Want to avoid overfitting and underfitting (head towards generality)
- Keep the test set separate at all costs
- Compare apples to apples
- One best performance metric does not equal best model

6. Experimentation

“How could we improve/what can we try next?”

Experimentation

- Try out a different approach for improving the machine learning model

Tools

Your computer
ANACONDA
jupyter
matplotlib
NumPy
pandas
scikit learn
dmlc
XGBoost

Tools mapping