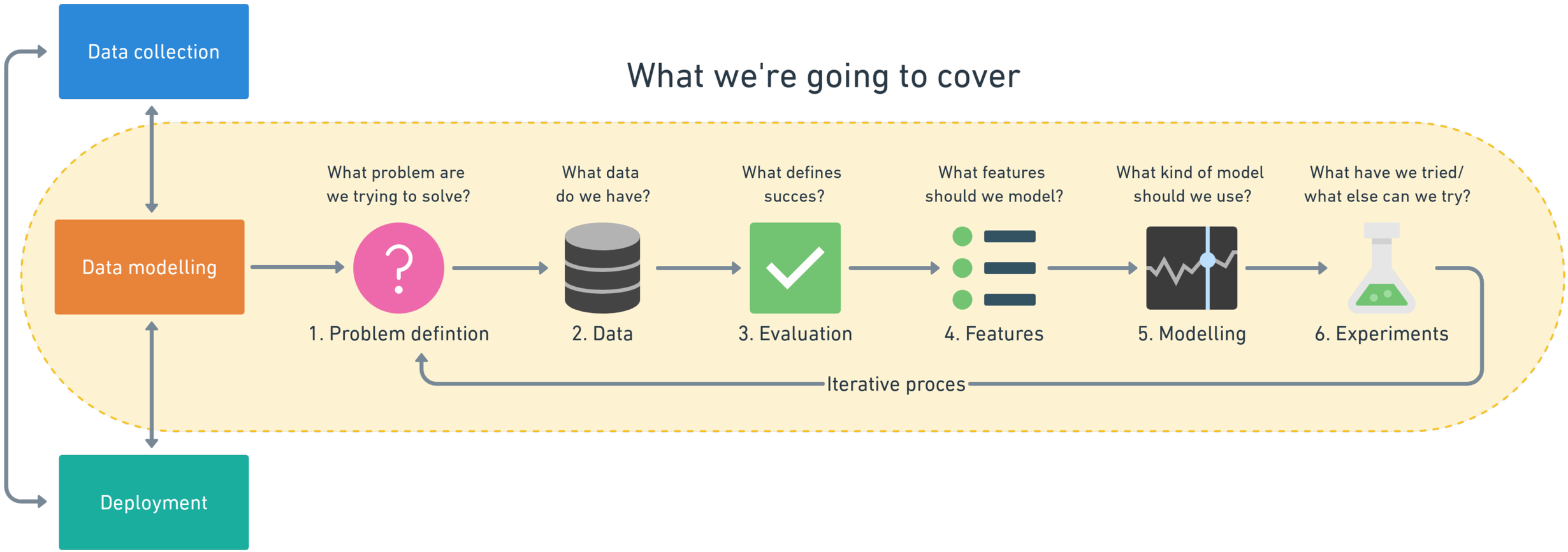
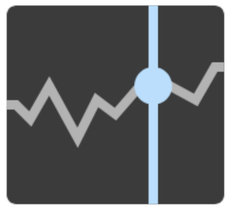
Steps in a full machine learning project



5. Modelling Part 2 — Choosing 

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

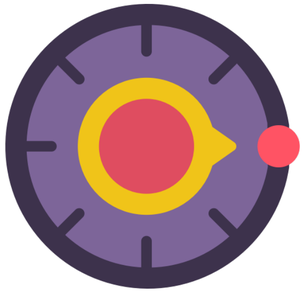
3 parts to modelling

1. Choosing and training a model

or

Training Data

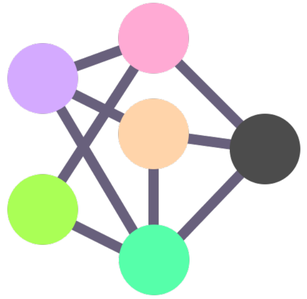
2. Tuning a model



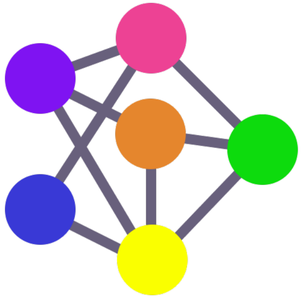
Validation Data

3. Model comparison

vs.



vs.



Test Data

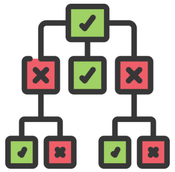
Choosing a model

Problem 1

Model 1

Structured Data

CatBoost



Random Forest

Problem 2

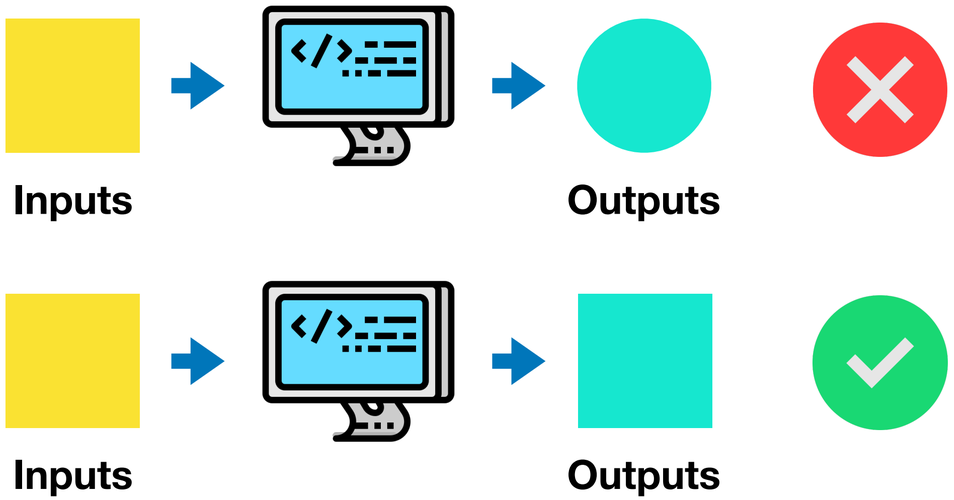
Model 2

Unstructured Data

Deep Learning

Transfer Learning

Training a model



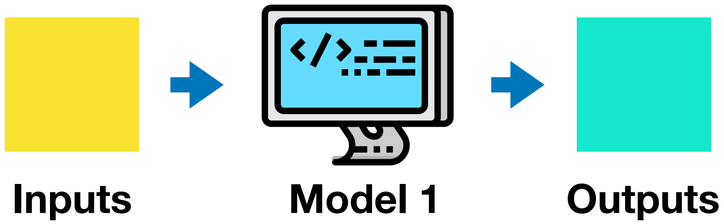
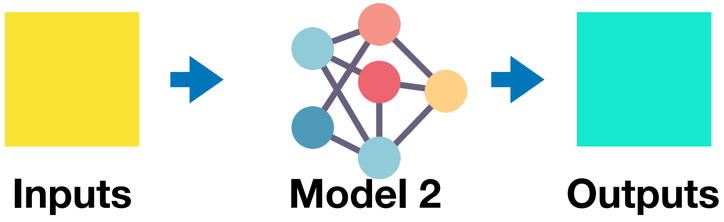
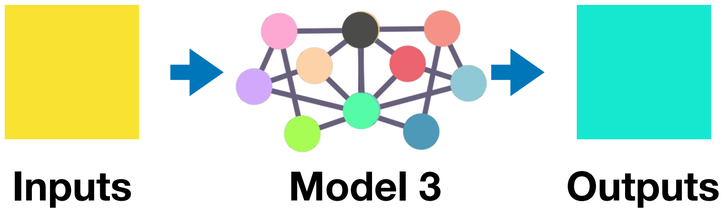
X (data) y (label)

ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?
4326	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: Minimise time between experiments

Experiment		Accuracy	Training time
1		87.5%	3 min
2		91.3%	92 min
3		94.7%	176 min

Things to 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

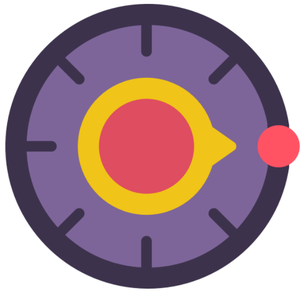
Up next

1. Choosing and training a model

or

Training Data

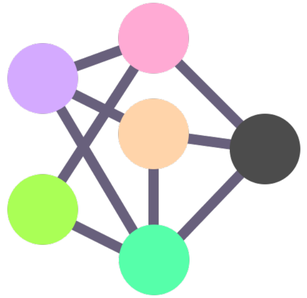
2. Tuning a model



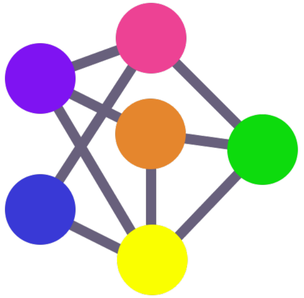
Validation Data

3. Model comparison

vs.



vs.



Test Data