KBTG x MeowCode

Say “Hello” Machine Learning

Jetsukda Janthornthao

AI Engineer @ Ztrus

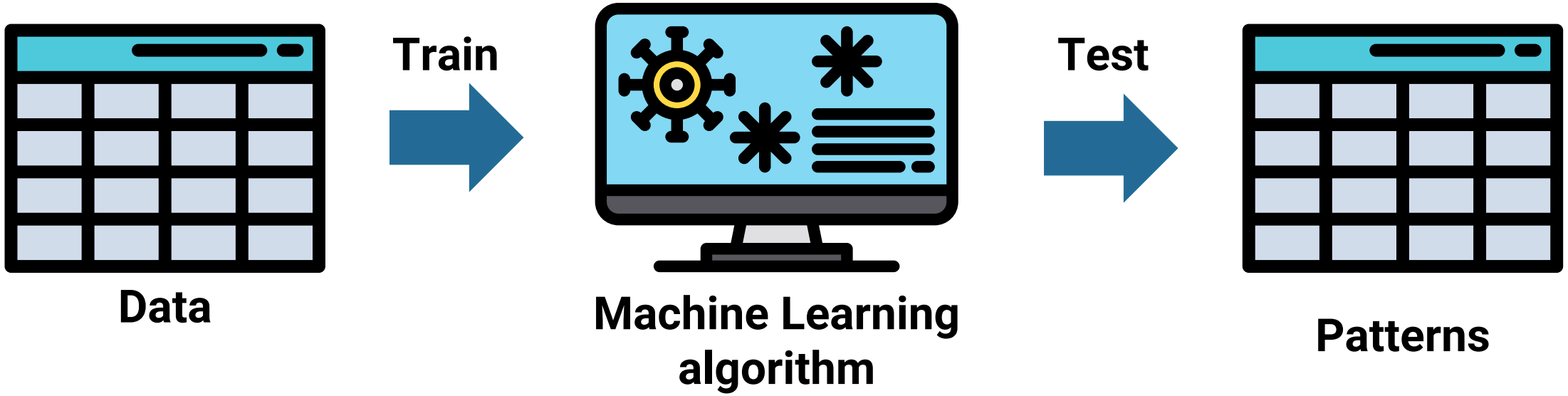
KBTG x 

Class Overview

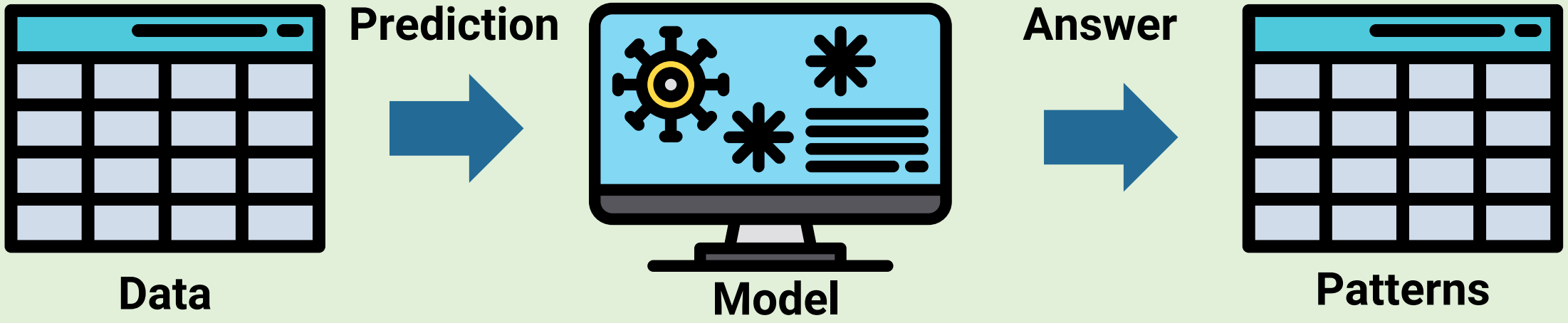
- What is Machine Learning (ML)
- Main types of Machine Learning
- Steps in a full ML project
- Machine Learning Framework
- Setup a Python Environment for Machine Learning

What is Machine Learning (ML)

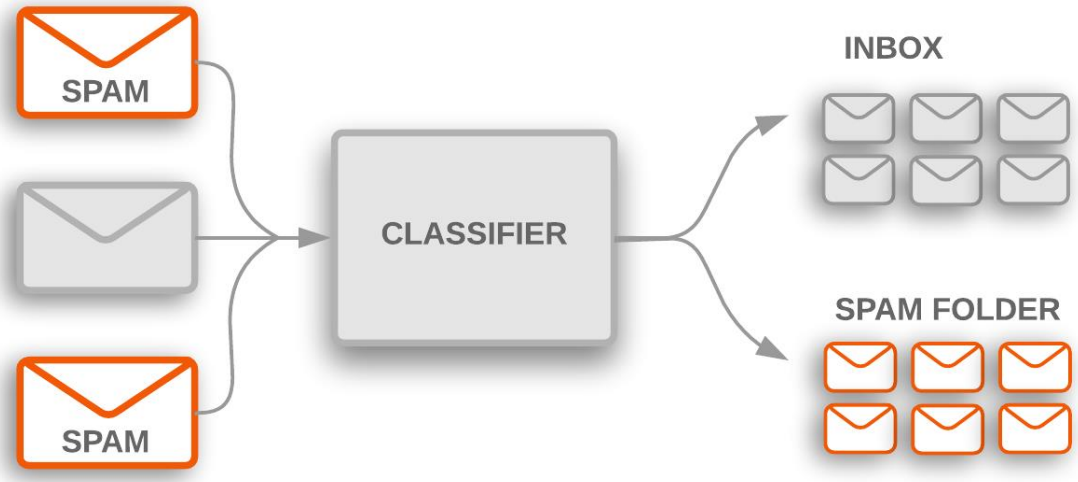
1. What is Machine Learning (ML)



Future

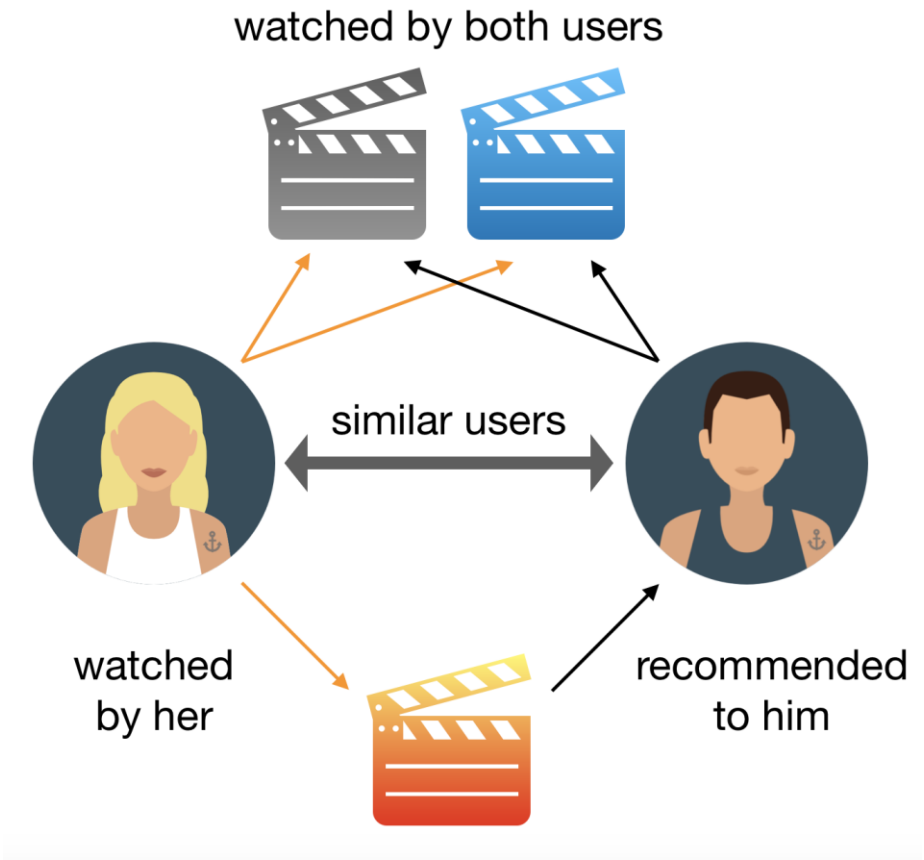


Spam Classification



- Email (text) as the input
  - Go into classification model
  - Output answer whether this is spam or not.

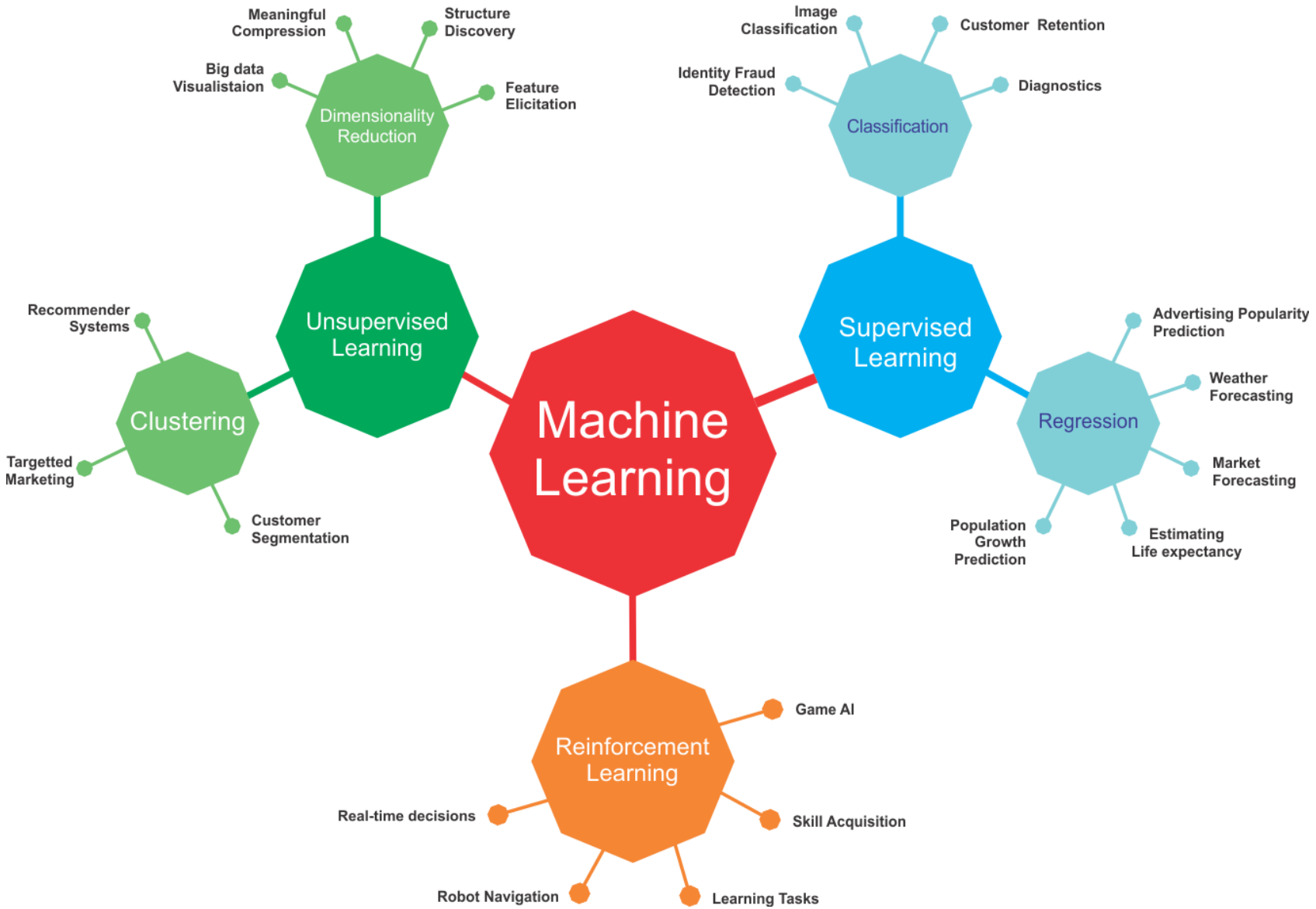
Recommended system



- Recommender systems aim to predict users' interests and recommend product items that quite likely are interesting for them.

Main types of Machine Learning

2. Main types of Machine Learning



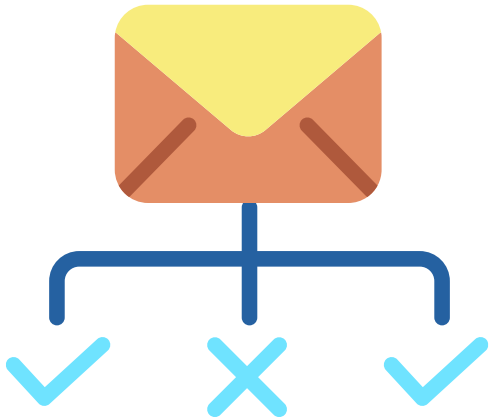
Terminology about ML

- Regression Problem
  - Predicting a continuous output (e.g. price, sales).
- Classification
  - Predicting a categorical output.
- Feature
  - With respect to a dataset, a feature represents an attribute and value combination.
- Label
  - In supervised learning, the "answer" or "result" portion of an example.

Supervised Learning

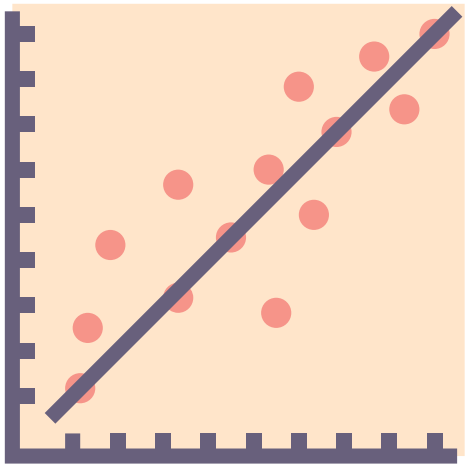
the training set you feed to the algorithm includes the desired solutions, called labels

Classification



categorical

Regression



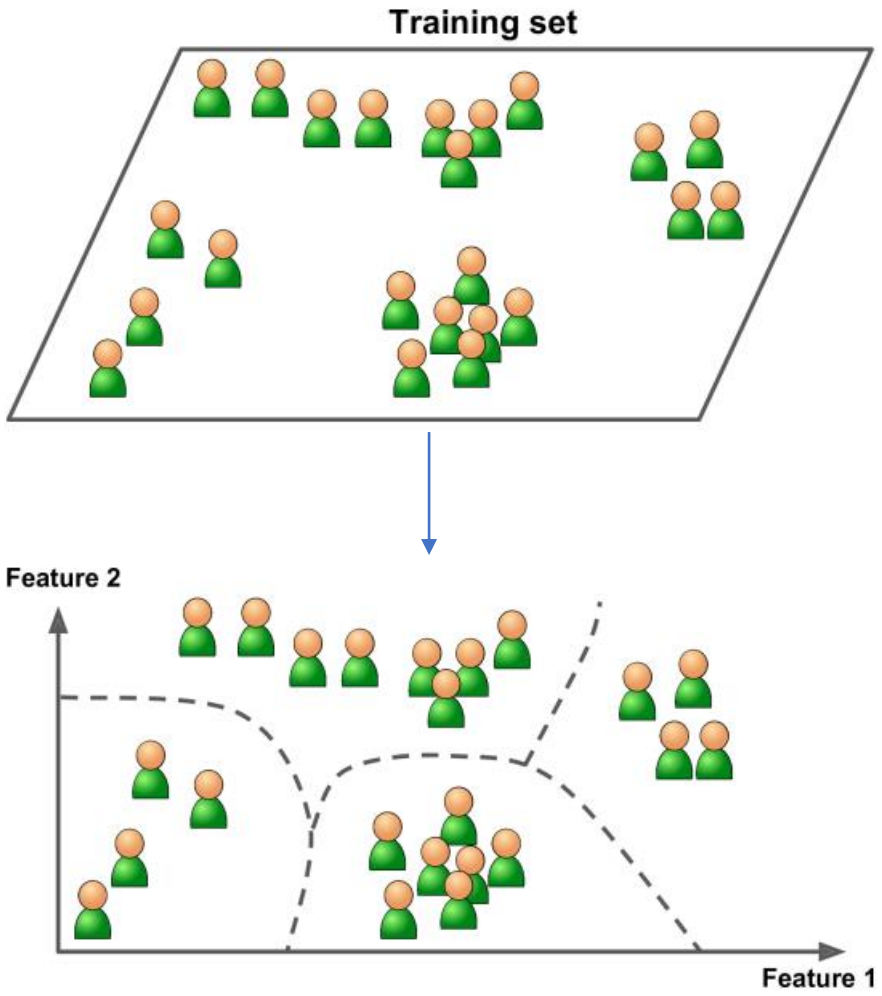
numerical

Supervised Learning Algorithms

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests

Unsupervised Learning

Training a model to find patterns in a dataset, typically an unlabeled dataset.

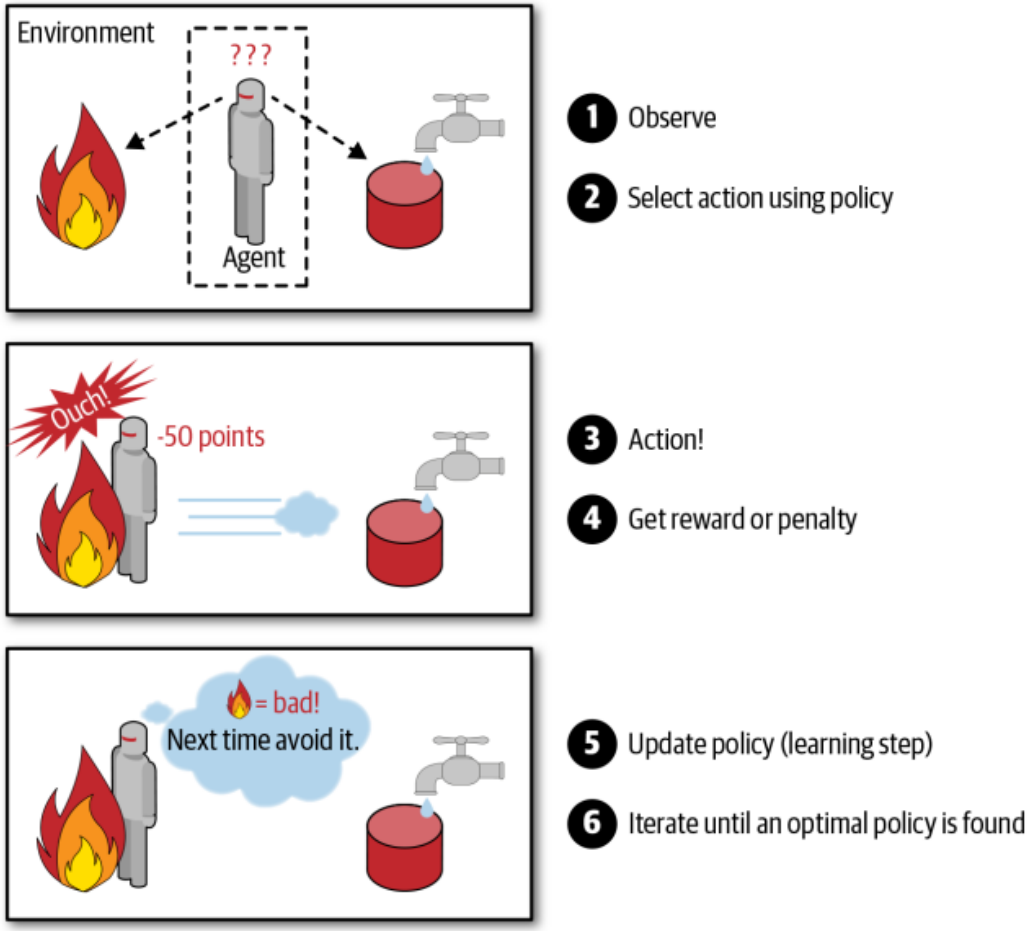


Unsupervised Learning Algorithms

- Clustering
  - K-Means
- Dimensionality Reduction
  - Principal Component Analysis (PCA)
- Anomaly detection
  - One-class SVM

Reinforcement Learning

The learning system, called an agent in this context, can observe the environment, select and perform actions, and get rewards in return

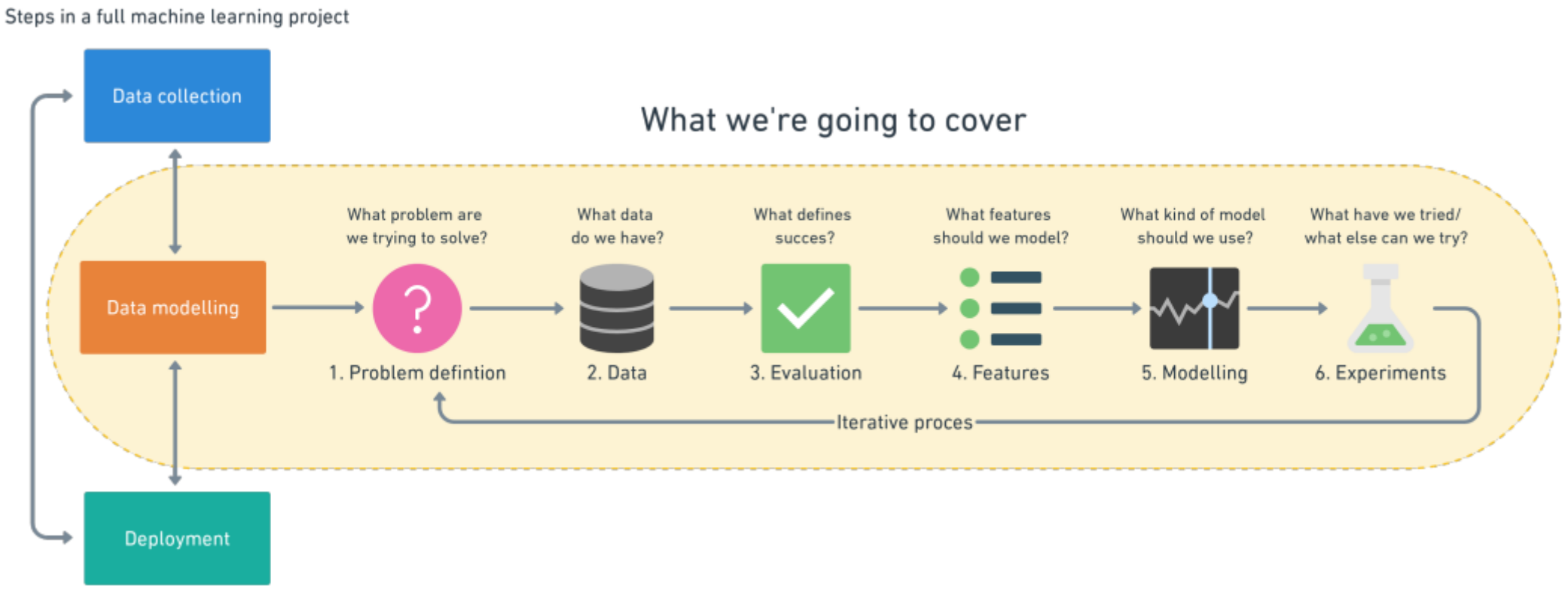


Reinforcement Learning Algorithms

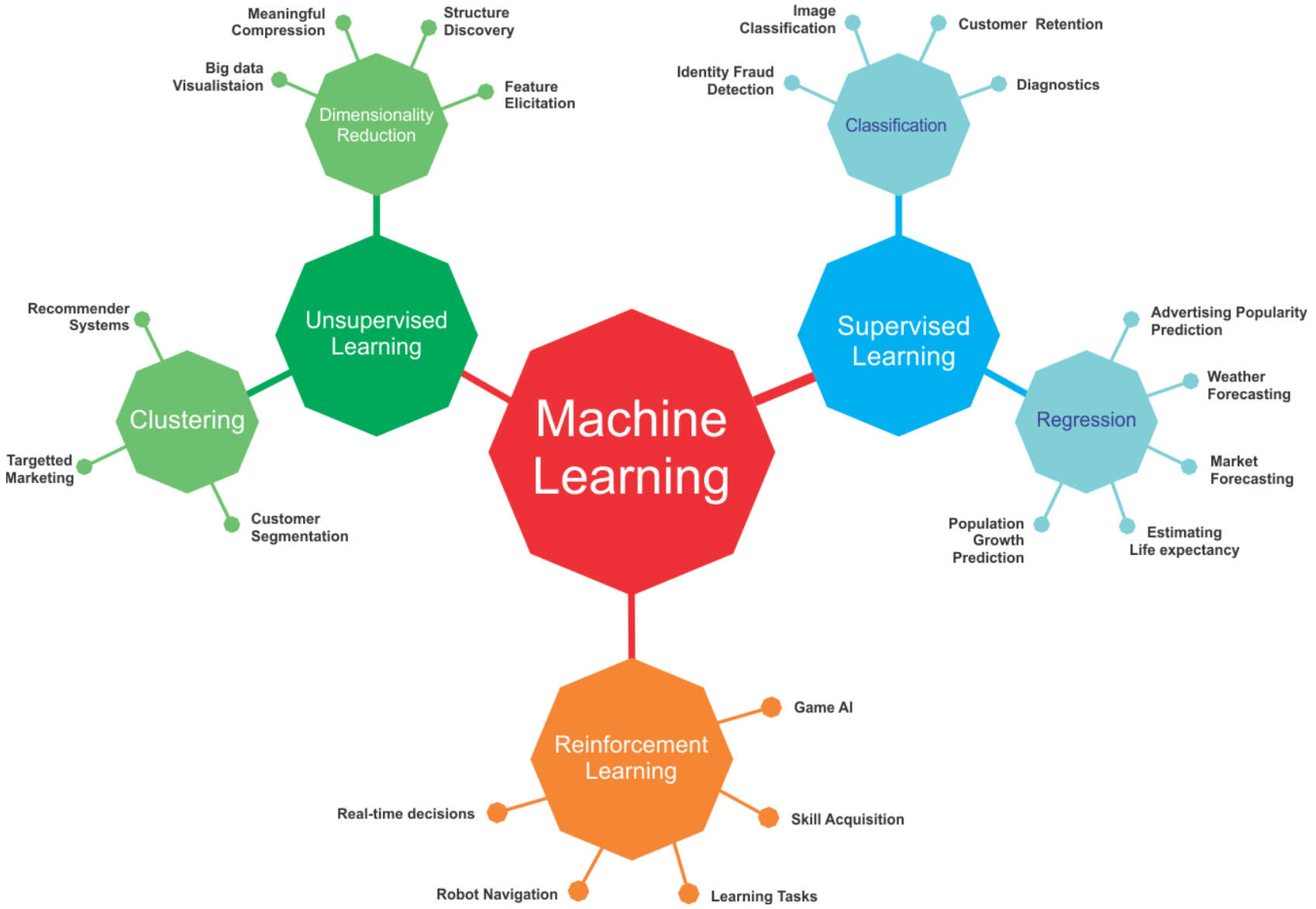
- Monte Carlo
- Q-learning

Steps in a full ML project

3. Steps in a full ML project

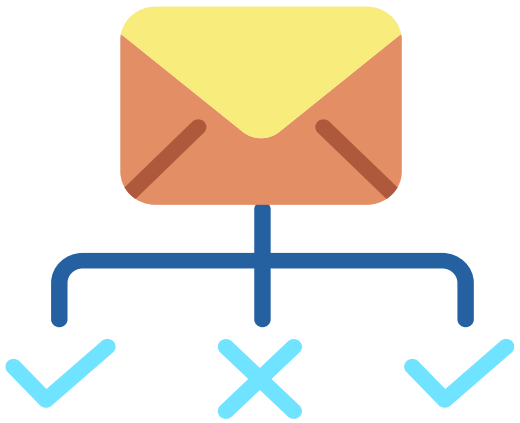


3.1 Problem definition



Supervised Learning

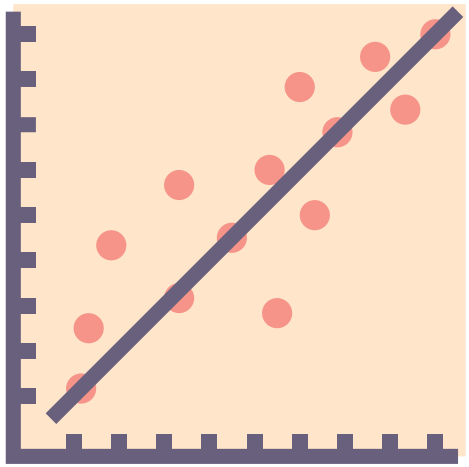
Classification



- Binary classification
  - Email spam detection (spam or not).
  - Churn prediction (churn or not).
- Multi-class classification
  - classify a set of images of fruits which may be oranges, apples, or pears.

Supervised Learning

Regression



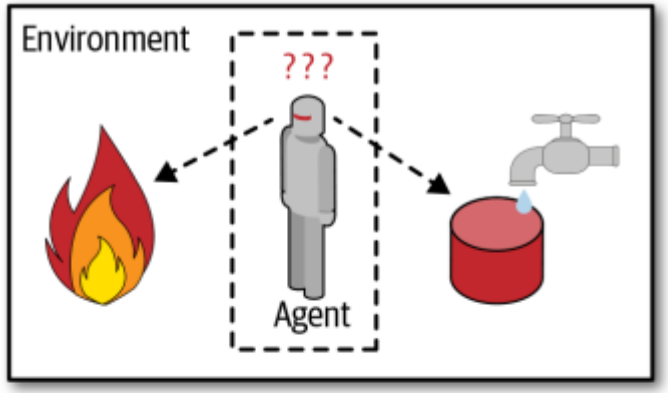
- Predicting whether stock price of a company will increase tomorrow ?
- How much will this house sell for ?
- How many people will buy this app ?

Unsupervised Learning: Clustering

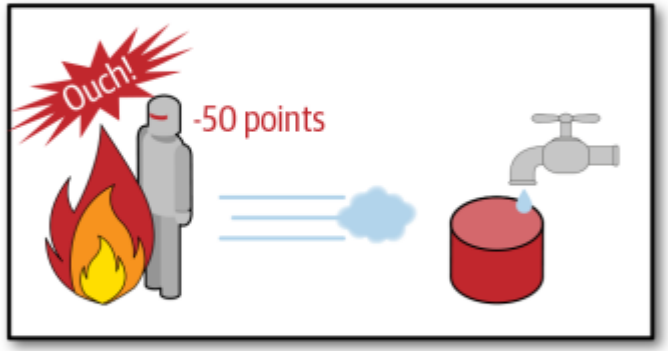
sample

Cluster/group

Reinforcement Learning



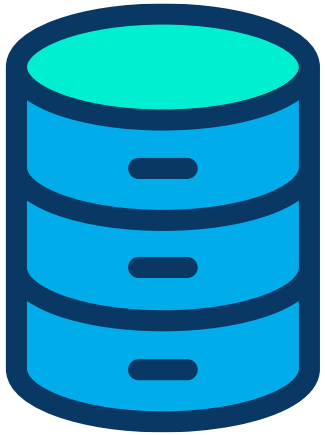
1. Observe
2. Select action using policy



3. Action!
4. Get reward or penalty

5. Update policy (learning step)
6. Iterate until an optimal policy is found

3.2 Data



What kind of data do we have ?

Different types of data

Rows

Columns

ID	Weight	Sex	Blood Pressure	Chest pain	Heart disease?
4326	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

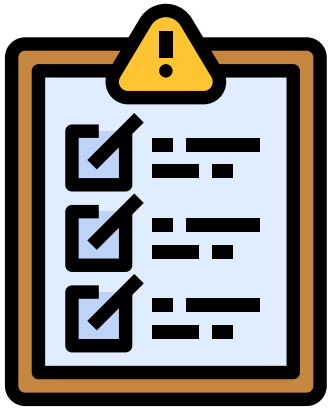
Structured

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

3.3 Evaluation

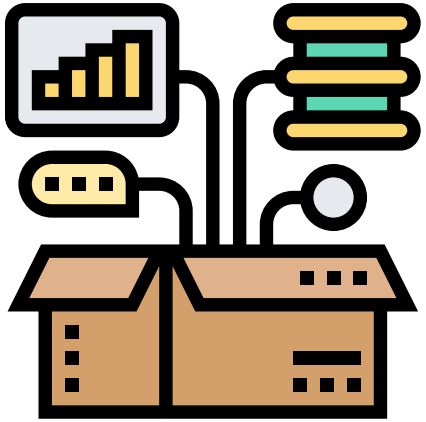


What defines success for us?

Different 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)	

3.4 Features



What do we already know about the data?

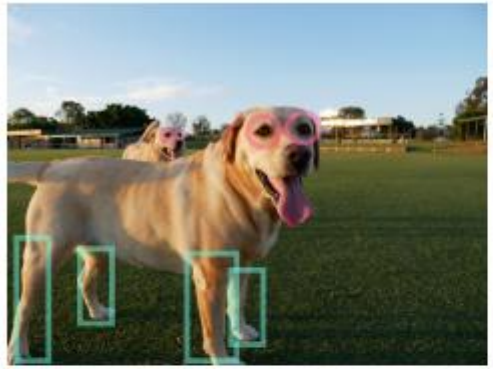
Different features of data

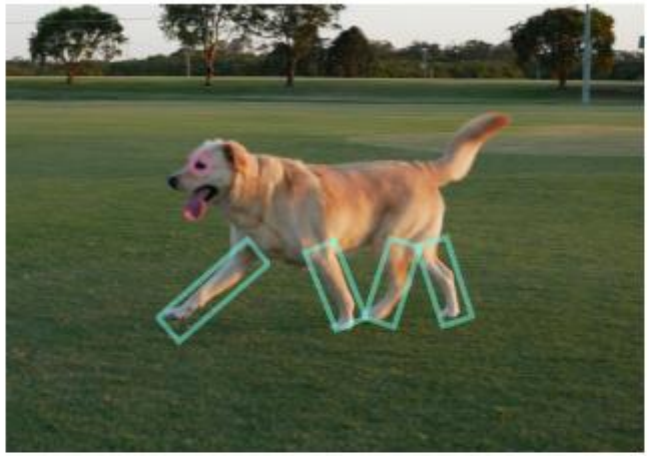
Country	GDP per capita (USD)	Life satisfaction
Hungary	12,240	4.9
Korea	27,195	5.8
France	37,675	6.5
Australia	50,962	7.3
United States	55,805	7.2

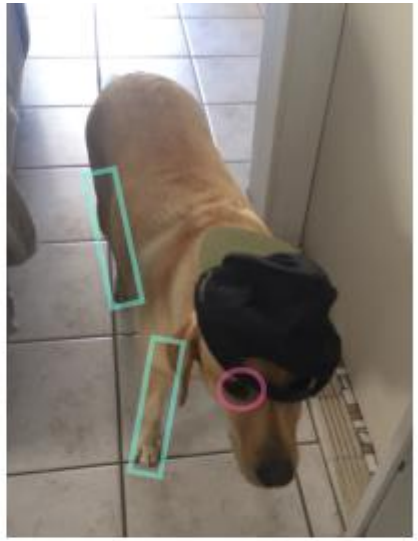
Categorical features

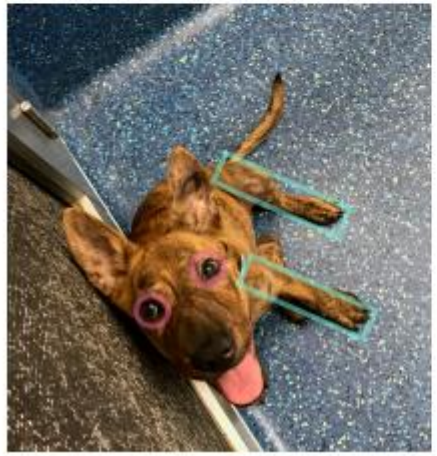
Numerical features

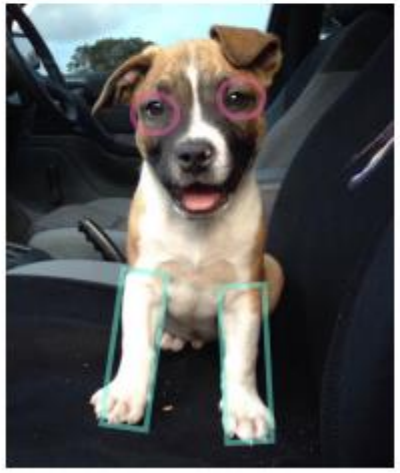
Different features of data



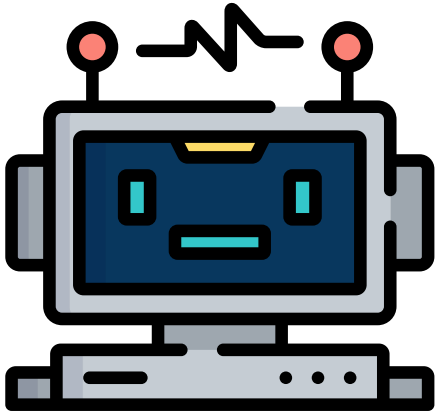








3.5 Modelling



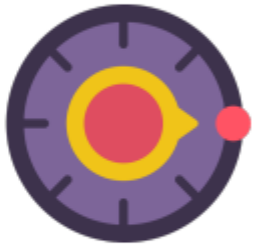
Based on our problem and data, what model should we use?

3 parts to modelling

1. Choosing and training a model

or

2. Tuning a model

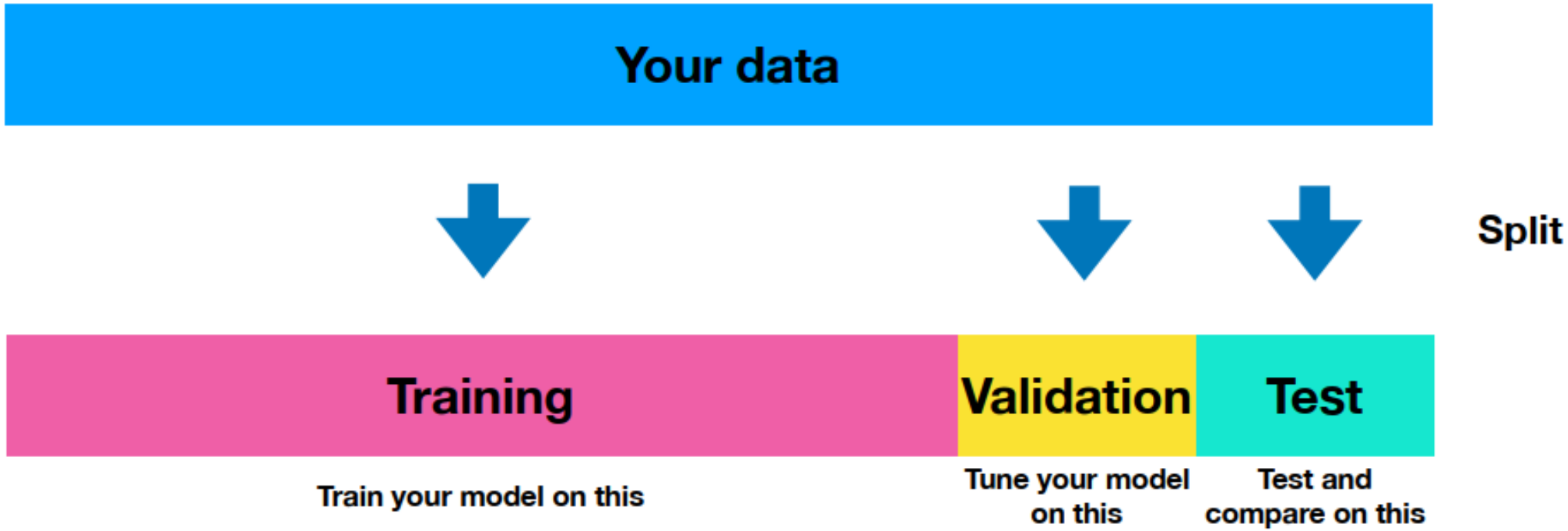


3. Model comparison

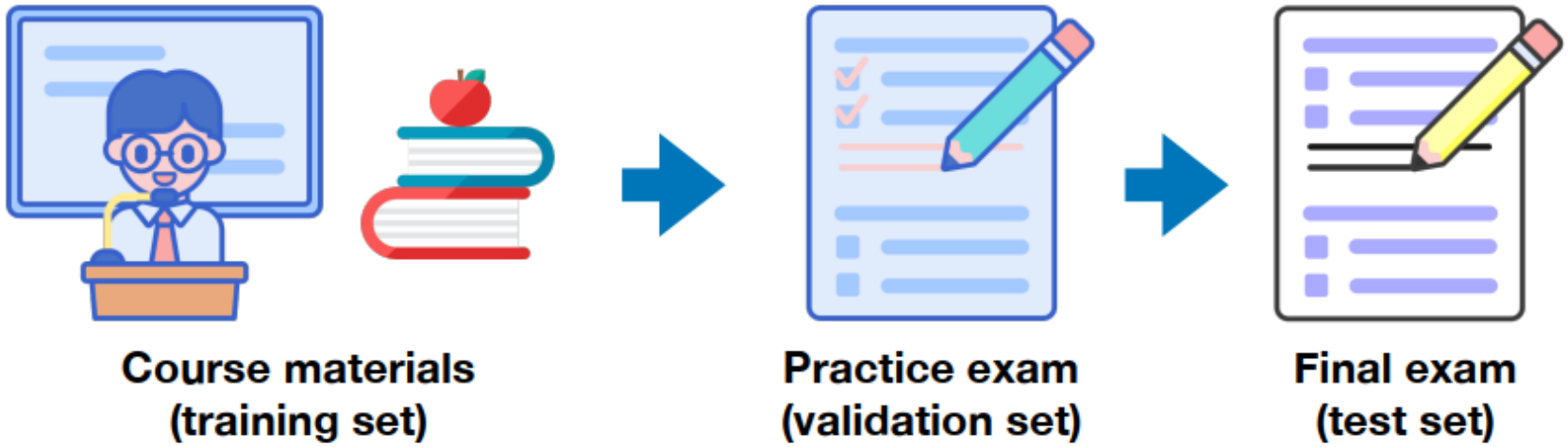
vs.

vs.

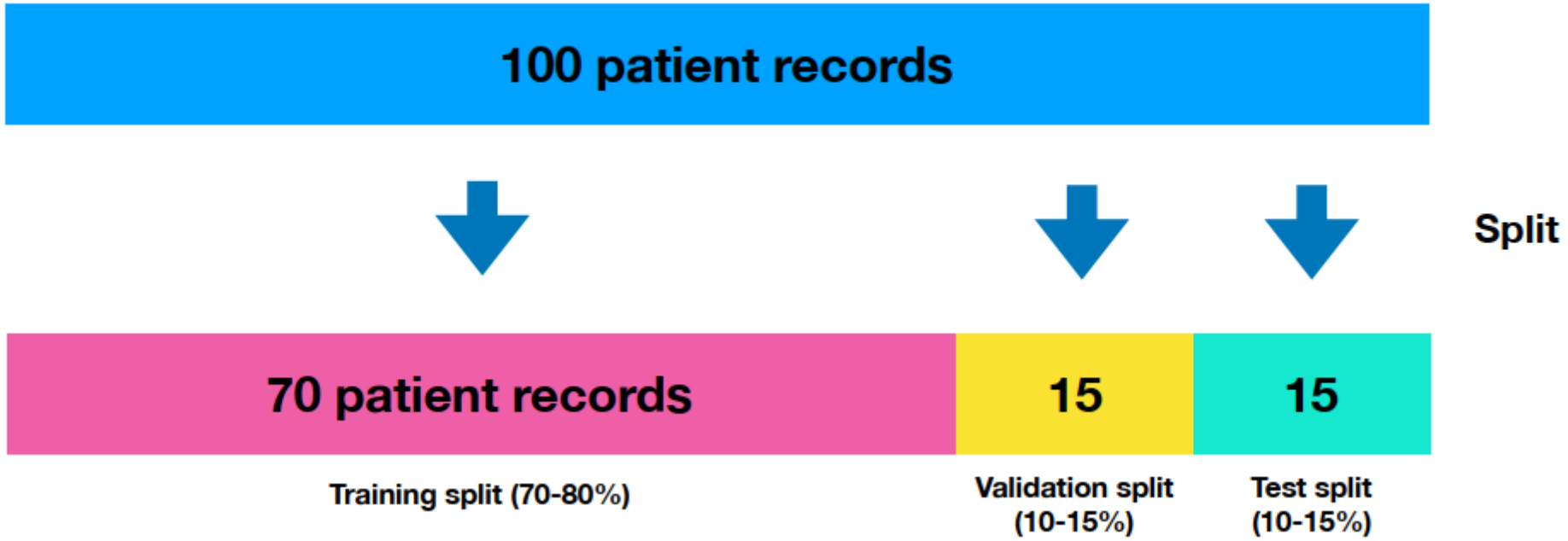
The most important concept in ML



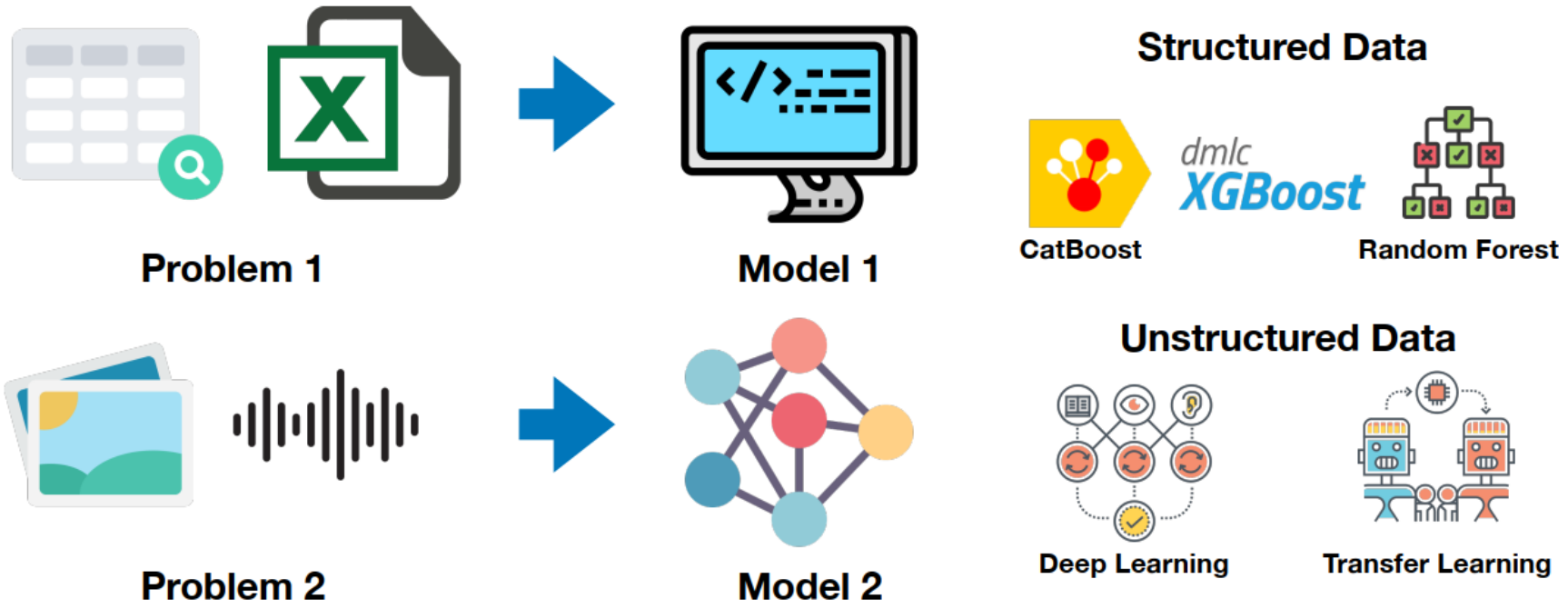
The most important concept in ML



The most important concept in ML

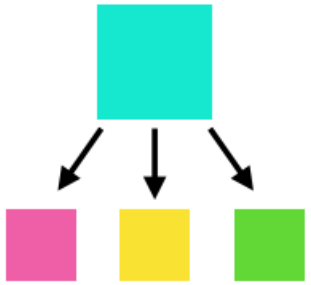


Choosing a model

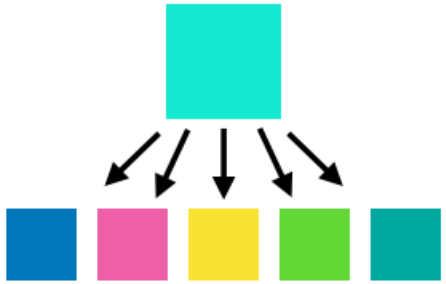


Tuning a model

Random Forest



3 trees



5 trees

Neural Networks

2 layers

3 layers

Testing a model

Data Set	Performance
Training	98%
Test	96%

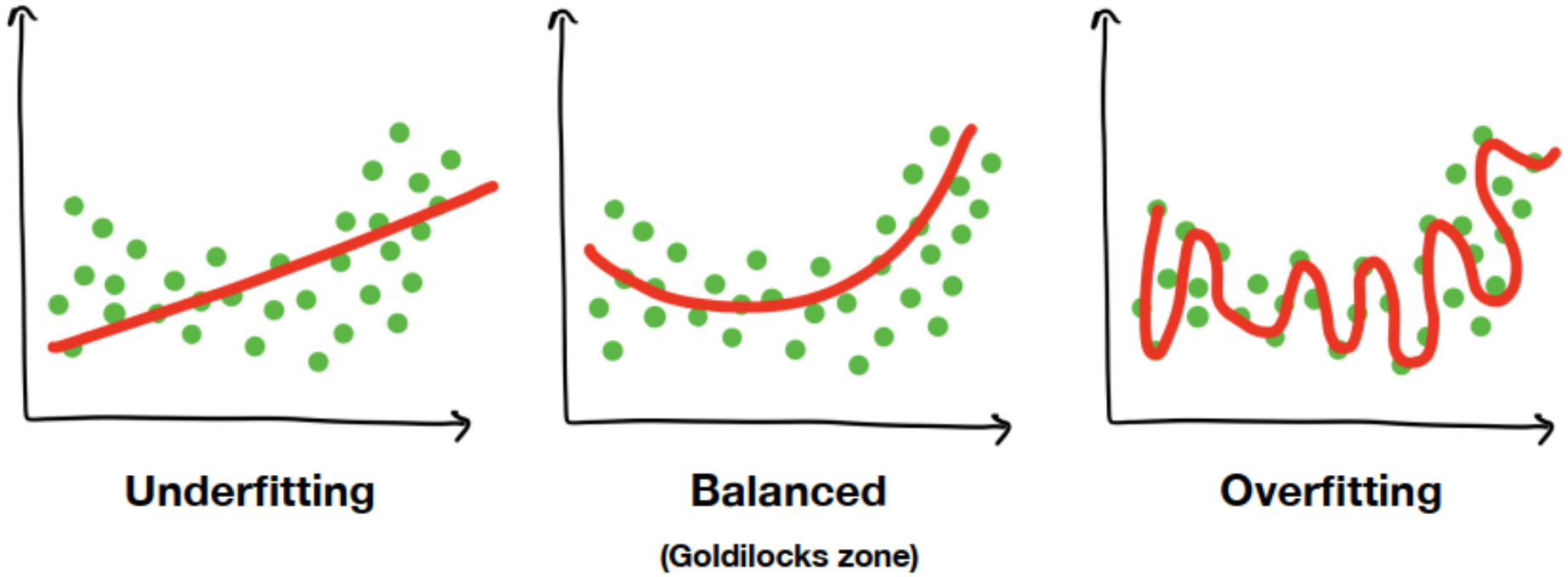
Underfitting (potential)

Data Set	Performance
Training	64%
Test	47%

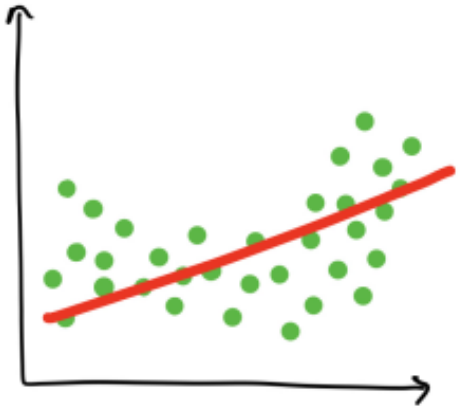
Overfitting (potential)

Data Set	Performance
Training	93%
Test	99%

Overfitting and underfitting

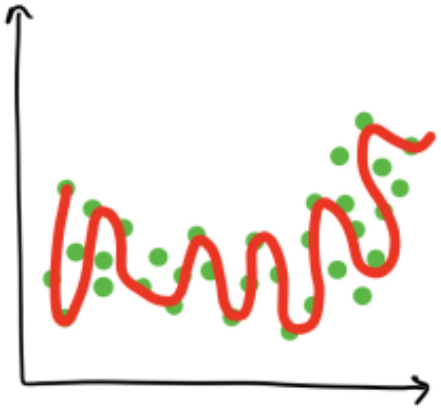


Fixes for overfitting and underfitting



Underfitting

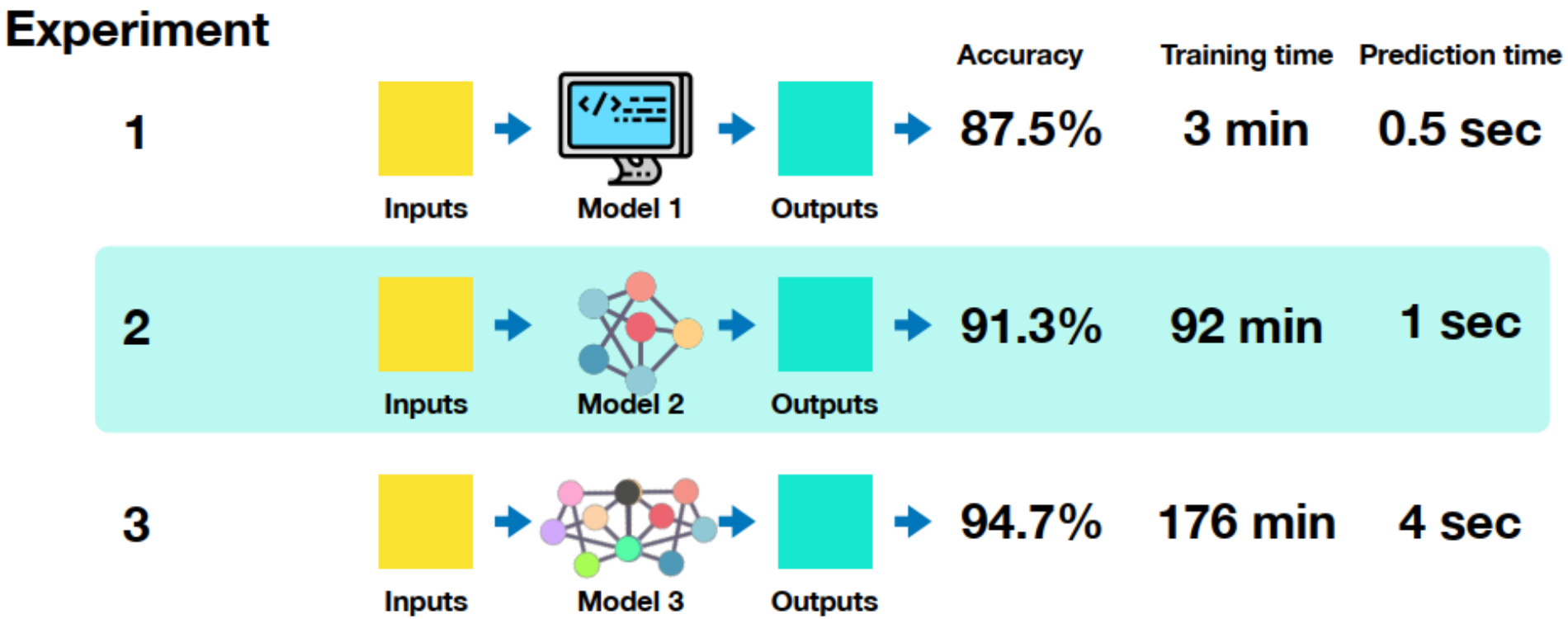
- Try a more advanced model
- Increase model hyperparameters
- Reduce amount of features
- Train longer



Overfitting

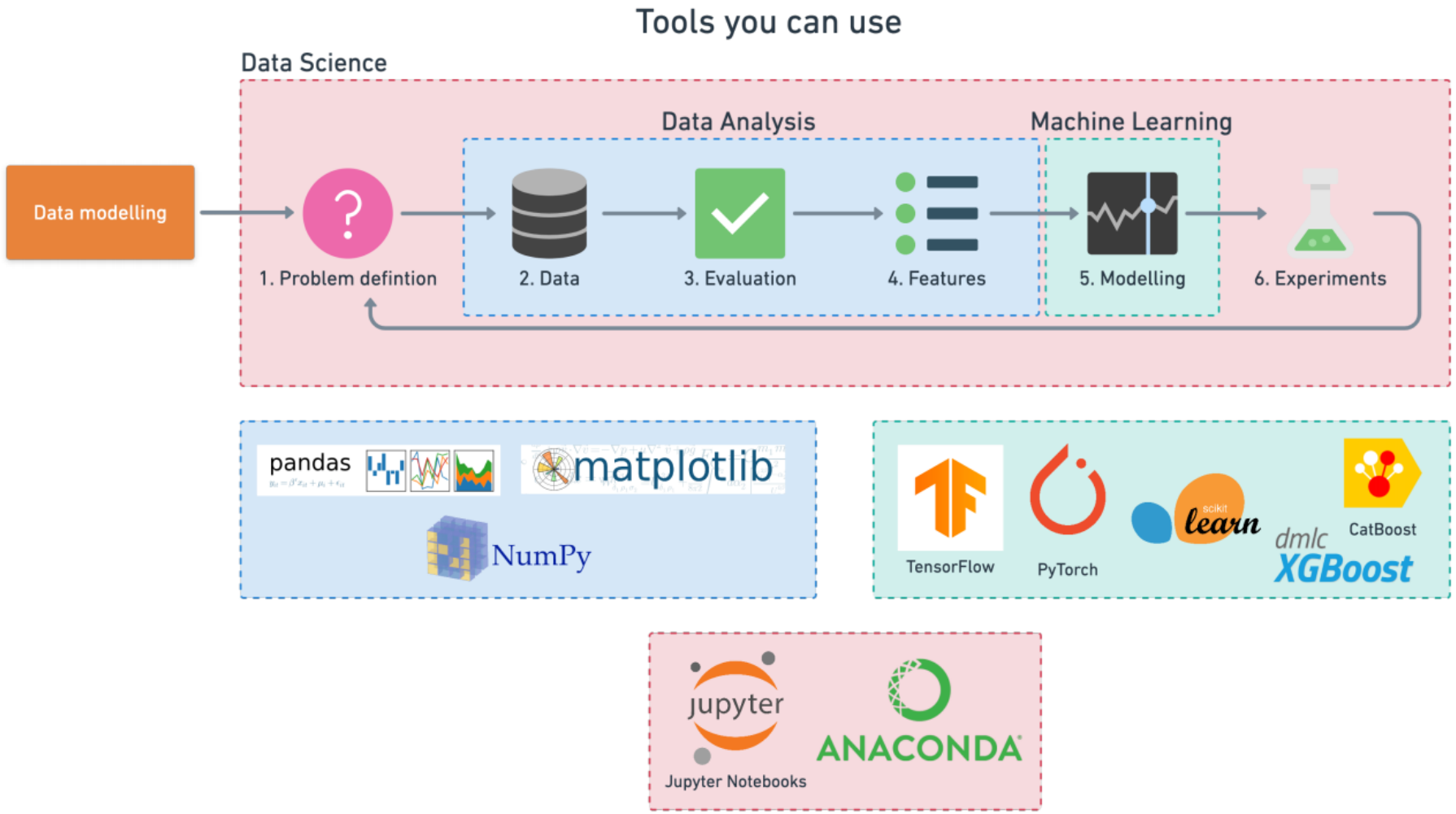
- Collect more data
- Try a less advanced model

3.6 Experiments



Machine Learning Framework

4. Machine Learning Framework



4. Machine Learning Framework

Reference

- Machine Learning Glossary
- zero-to-mastery-ml
- machinelearning-datascience
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow
- handson-ml2
- scikit-learn: Save and Restore Models
- Introduction to Machine Learning Algorithms: Linear Regression