Machine learning

IB DP Computer Science A4 Machine learning full package

A program that learns from examples

> Patterns from data, not rules from a programmer.

- Theme A4: what ML is, its types, how models are judged, and ethics.
- Prerequisites: scatter graphs and percentages.

AI, ML and deep learning

Deep learning is part of ML, which is part of AI.

- AI: computers doing tasks that seem to need human intelligence.
- Deep learning uses neural networks with many layers.

Three types of learning

Type	Learns from	Example
supervised	labelled data	spam or not; a price
unsupervised	unlabelled data	grouping shoppers
reinforcement	rewards and penalties	a robot learning to walk

Training and testing

> Test on data the model has never seen.

- **Overfitting**: memorises the **training** data, fails on new data.
- Underfitting: too simple to capture the pattern.

When 90% accuracy is useless

> Always saying "healthy" scores 90% and finds no one.

- Accuracy = correct / total x 100%: 90 of 100 is 90%.
- Check how the model does on each class.

Hardware for ML

Training

Thousands of cores in parallel.

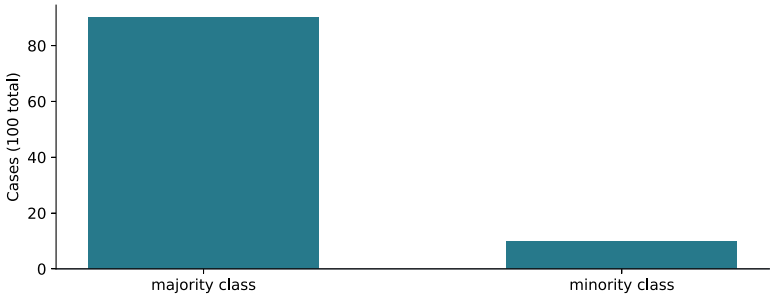
GPUs or special chips, often in the cloud.

Inference

Using the trained model.

Far less power - it can run on a phone.

Imbalanced classes

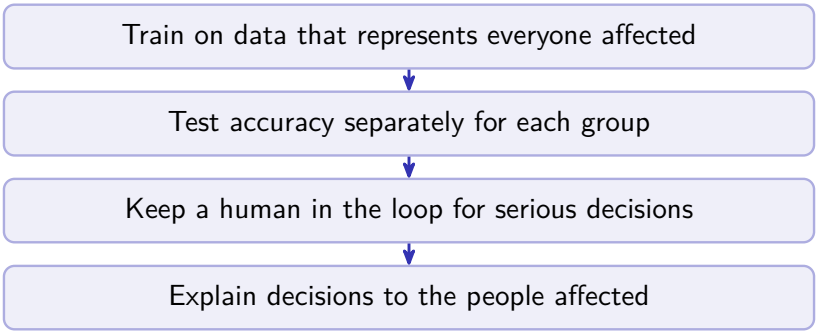


Accuracy alone can mislead

Who is affected?

Issue	What goes wrong
bias	face recognition less accurate for darker skin
transparency	a loan refused with no reason given
accountability	who answers for a self-driving crash?
privacy	photos scraped without consent

Reducing the harm



Practice

Your turn

Try these:

A. Name the type: grouping songs with no labels.
B. Why test on data the model was not trained on?
C. 72 correct out of 80: the accuracy?
D. Two ethical issues of shortlisting applicants by ML.

Check yourself

