# Machine Learning

## Introduction

Machine Learning is often considered a buzzword, whose use has been overstated (especially in engineering where deterministic, explainable models & results are key).

It may have been overestimated and misunderstood by some, but this series of posters aims to give you a high level understanding of machine learning such that we can objectively assess its relevance to our work as a department.

There'll be a poster on a new topic every 2-3 weeks. Each will consist of an A3 sheet, the front explaining the concept at a high level and the back in more depth, showing how you can implement it yourself using Python!

## Installing Python

If you don't have Python installed and would like to try these models yourself, you can download and install Anaconda from PC services:

1. Open PC services
2. Type "anaconda" in the "Search for a software" bar and search
3. Click "Anaconda for Python 2.7" (3.6 is fine, but will be coding in 2.7 here)
4. Click Download
5. Follow the installation instructions
6. Once you've opened anaconda, launch jupyter notebook (instructions on how to use can be found at https://jupyter-notebook.readthedocs.io/en/stable/)

Source code and data for the examples will be uploaded to P:\ENGINEERING\FPO\Digitalisation\1 FPO Projects\IIX Data Analytics\02-General presentations\Knowledge folder - preparation\Data Science Learning Framework\Machine Learning Posters, and you can use them to write your own models if you'd like.

## Definitions

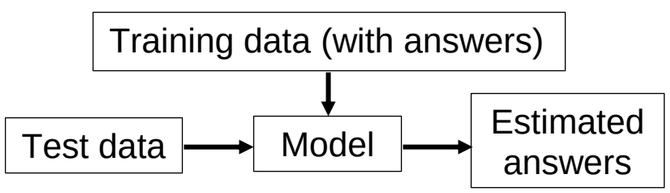
"A computer program is said to learn to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with E."[1]

This is a common definition of machine learning, the key part of which is that it's learning from *experience*. The implication here is that the program is not being explicitly programmed to improve performance.

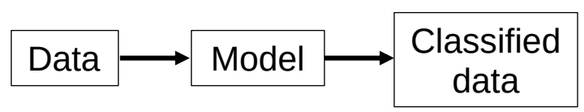
There are several other important definitions to be aware of:

- Training and test data:
  - Training data is used to tune the model to give better predictions.
  - Test data is used to validate the model.
- There are two main types of model, supervised and unsupervised learning:
  - Supervised learning: the model is given the "right answers" to train itself.
  - Unsupervised learning: the model isn't given any answers, and may draw its own conclusions.
  - Other types such as reinforcement learning, recommender systems exist.

**Supervised Model**




**Unsupervised Model**




- The two main types of problem are regression and classification:
  - Regression: Attempting to predict continuous value outputs.
  - Classification: Discrete answers, attempting to sort data into groups.

[1] – Mitchell, T. (1997) *Machine Learning*. McGraw Gill