[dstl]

The Dstl Biscuit Book

Artificial Intelligence, Data Science and (mostly) Machine Learning

1st edition revised v1_2

Foreword

In recent years Artificial Intelligence (AI), Data Science and Machine Learning have rapidly risen to prominence, offering breakthroughs in many areas previously beyond the capabilities of traditional approaches to computing. They offer the ability for machines to provide new insight into the vast quantities of data that are produced at an ever-greater rate and sophisticated actions to be performed that previously could only be accomplished by humans. Such ability has become possible due to a combination of the volume of data produced and increased computational capability. In particular, these advances have been made possible through the development of Machine Learning methods; so while this guide will cover Artificial Intelligence and Data Science, mostly it is about Machine Learning.

In the wake of this revolution, there is left a trail of bewildering terminology and hype. This short guide is intended to provide some explanation of the most common terminology and help you separate fact from fiction.

Defence Science and Technology Laboratory (Dstl)

Dstl delivers high-impact science and technology for the UK's defence, security and prosperity.

Our world-leading AI, Data Science and Machine Learning work involves working in partnership with specialists from industry, academia and allied nations. We look at everything, from very-early research looking at how machines interact with humans, to applying data science to real-world challenges and operational requirements. This is the future, not only of defence and security, but of the world we live in.

In our role, we are able to provide clear guidance on the engagement with and use of AI, Data Science and Machine Learning in a defence & security environment.

Dr Mercedes Torres Torres[1], Glen Hart[2] and Toni Emery[2]

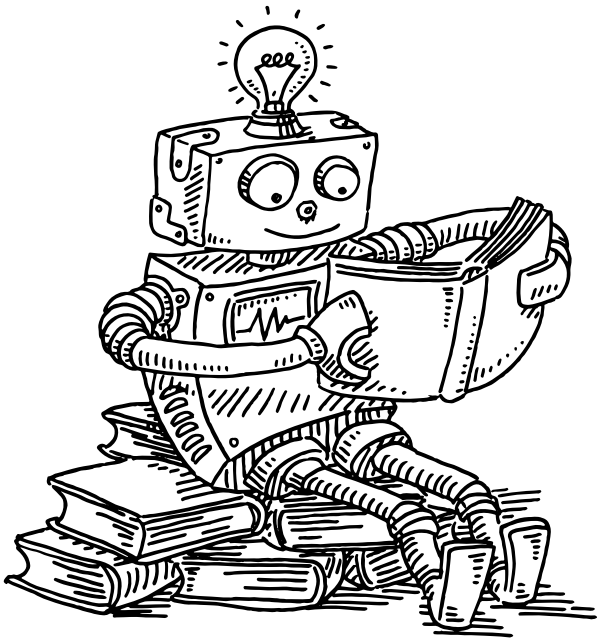
The Biscuit Book has been developed by Dstl with input from the AI Micropedia © Dr Mercedes Torres Torres 2019, with the kind permission of the author.

1 University of Nottingham
2 Dstl

Introduction

This guide is what we call a Biscuit Book, something you can pick-up and dip into with a tea and biscuit. The Biscuit Book is arranged as a series of easily digestible chunks that each cover a topic and doing so in a manner that provides the essential information without ever being too technical.

We hope you find the Biscuit Book both informative and digestible, although we do not suggest dunking it in your tea!



Definitions

So what are AI, Data Science and Machine Learning?

There are no universally accepted definitions for Artificial Intelligence and Data Science; Machine Learning is generally better defined.

It is not unusual to see all three used interchangeably in industrial, commercial, or non-expert settings. Many product descriptions, companies and media outlets use these terms in a very loose fashion.

A recent study analysed over 2,800 European start-ups claiming to use AI and found that only 40% of these actually did. This should alert you to the fact that there are lots of different definitions and there is potential for a lot of argument about what AI, Data Science and Machine Learning are, or are not.

Given the confusion, some clarity is necessary. For this reason, we provide the following simple definitions to give you some sense of how they differ from each other:

Artificial Intelligence

Theories and techniques developed to allow computer systems to perform tasks normally requiring human or biological intelligence.

(As you'll see later the intelligence is very limited.)

Data Science

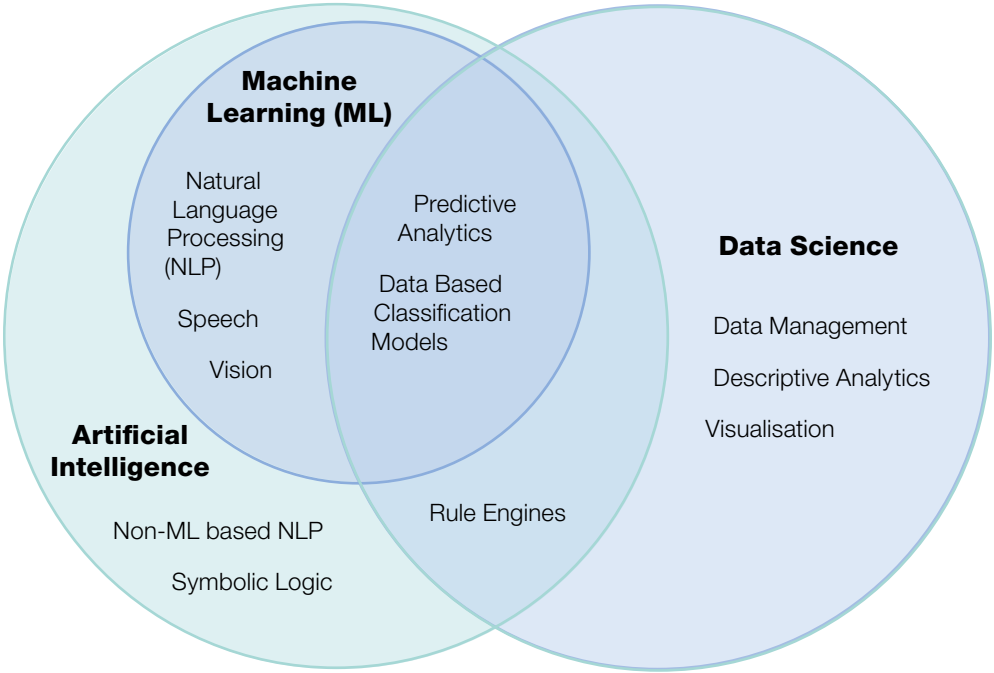
A multidisciplinary field that combines statistics, mathematics, computer science and domain expertise to extract relevant insights or knowledge from data.

Machine Learning

A field that aims to provide computer systems with the ability to learn and improve automatically without having to be explicitly programmed.

So AI is about performing tasks intelligently, Data Science is about discovering insights from data, and Machine Learning is a means to achieve both through automatic processes. It is easy to see how terms can be muddled: you may ask for AI and get Machine Learning because Machine Learning is the method that is applied to achieve the AI.

The diagram below shows all three topics in relation to each other and some applications and techniques that are applicable to them.



Examples

Here are some examples of daily activities that you might be doing and that rely on Artificial Intelligence, Data Science and Machine Learning:

Search engines:

Google, Bing, and other search engines use sophisticated machine learning methods to find and rank webpages that match your search criteria. These engines not only use machine learning to provide relevant results for you, they also combine data science and machine learning so every time you search for something, the algorithms at the backend will monitor your responses: which pages do you open, how many do you open, how long you stay in each, etc. This way, these engines can tailor the search results to you.

Virtual Personal Assistants:

Have you used Alexa, Siri or Google Home? All of these virtual personal assistants apply data science to complete tasks such as answering simple questions, telling you the news or weather, or playing music or podcasts. To do this they collect information about what you are saying, and also about when, where and how you are saying things. These assistants then use this information to produce results that are tailored to your preferences. They also use machine learning to: 1) understand you (speech processing and understanding); 2) improve their performance based on your previous interactions; and 3) communicate back to you (dialogue management).

Traffic Status:

Ever wondered how a traffic or map app can tell you which section of your commute will have heavy traffic? That is because they are using the GPS location and speed of their users, then adding it to a central server managing traffic. Data Science methods are then used to build maps of current traffic and to estimate the density of the traffic. For areas in which GPS information might not be available, Machine Learning can be used to predict regions with heavy traffic using historical data.

Loan approvals:

Banks and other financial entities collect extensive information about customers who are applying for loans. Data Science is used to find relevant data, while machine learning is used to classify the customer as eligible or not for a loan depending on their history and the history of people with a similar profile.

Activity trackers:

Physical activity trackers, such as Fitbit, collect a vast array of information about their users. Data collected includes: steps covered, floors climbed, calories burned, sleep stages, heart rate per minute, etc. Data Science is then used to create health stats which, if the user allows, may be shared with external partners (such as health professionals and insurance companies) so they can provide a better and more personalised service.

Chatbots (Online Customer Support):

More and more websites now provide customer support using chats, but quite often the person you are chatting to is a chatbot not a person. Businesses like IKEA, Hotels.com, and E.ON use bots to filter any customers who might need to contact them. These bots use machine learning to identify relevant information in your text and provide possible answers to your queries. If the bots are not able to provide the information customers need, then they are transferred to a human representative. Duolingo, an app for learning new languages, uses chatbots to help users practice their newly-learned language skills via text messages. They also use Data Science to collect information about their users and apply machine learning to classify their personalities and learning styles, with the idea of allocating them to a chatbot that best matches them.

Recommendation systems:

Have you ever received an e-mail from Amazon with products that could interest you? Or have you ever seen the “Recommended for You” section on Netflix? These are two examples of recommendation systems. These systems collect and pre-process data from your activity within their site (i.e. what you search for, what you look at, for how long, what you place in your wish-lists or your cart, which parts of a movie you rewind or fast forward, etc.) to produce recommendations based on how your behaviour compares to the rest of users on the site. Using data science, they are able to group customers according to behaviour and share recommendations amongst each group. So, if several people with behaviours similar to yours have watched a movie that you have not, Netflix will recommend it to you.

And of course there are plenty of applications in a professional context including:

- Classification: for example, classifying images as containing vehicles, people etc.
- Recognition: a common application is facial recognition.
- Filtering: taking large volumes of images, video or documents and selecting those that contain certain images, objects or references.
- Anomaly detection: for example, analysing large quantities of engine performance data and identifying possible anomalies that could indicate a fault.
- Prediction: for example, where we may want to predict when the food is likely to go bad.

And so on, the range of applications is growing so fast this list could go on and on.

Artificial Intelligence

AI can be broadly divided into two categories: Narrow Intelligence and General Intelligence.

Narrow Intelligence, also known as Weak AI, exists, whereas General or Strong Intelligence is something machines could only dream about – if they had it!

Narrow Intelligence

Narrow Intelligence is AI that is focused on performing one main task. All AI Systems that currently exist have narrow AI although they may appear to be smarter than they are. An example of this is Alexa from Amazon which has a limited pre-defined range of operations that it can carry out; it does not possess any intelligence or self-awareness capabilities although we may have the illusion that it does.

General Intelligence

General Intelligence makes reference to machines that can perform many tasks, be cognitively aware of what they are doing and be able to self-learn and adapt. Examples of General Intelligence include HAL from 2001: A Space Odyssey, R2-D2 from Star Wars and Data from Star Trek. That all these examples come from Science Fiction should give you a clue that General Intelligence is a long-term aspiration (at least for some) rather than anything you will be able to put into use anytime soon. Perhaps because of the prevalence of General Intelligence in Science Fiction we often equate AI with 'human like' intelligence. In fact, there is no reason to suppose that this is the case and indeed there are many indications that machine intelligence will be quite different to human intelligence.

Further references to AI throughout this guide will refer to Narrow Intelligence.

AI Methods

AI isn't a single thing, but a collection of different methods that aim to meet the general objective of performing actions using human or animal-like intelligence. Even which methods are, or are not, AI is disputed. Here are summarised methods that are often included within the AI family.

Symbolic AI

Up until the late 1990s, AI was dominated by an approach now called Symbolic AI. Back then it was just AI, a time when both AI and the definition of AI were simpler. Symbolic AI is based on reasoning, typically through the application of a branch of Mathematics called First-order Logic. This approach was successful in certain areas such as Expert Systems – systems using rules to provide advice and guidance – but suffered from many limitations. It gave much promise but often great disappointment too.

Symbolic AI uses simple statements to provide basic knowledge. For example:

A is within B.

B is within C.

From this a Reasoner, a programme capable of making inference from such statements based on Logic, could infer that

A is also within C.

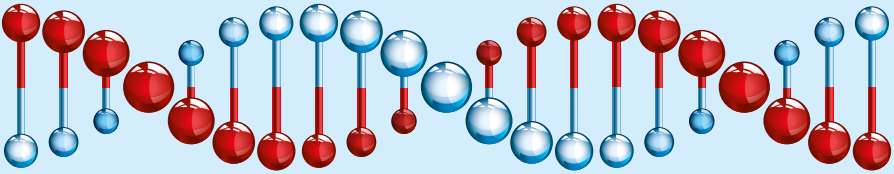
One problem with Symbolic AI is that it is built on rigid lines and struggles to match a human's ability to deal flexibly with the complex ways we understand the world. For example, if asked to think about a bird most people will think of something small and feathered, which flies and makes tweet tweet sounds. None of this is true of all birds but people don't struggle with this. It is however a major problem for Symbolic AI.

Symbolic AI is by no means dead and gone, it is still used today in many areas. One common use is in the development of Ontologies – formal descriptions of topic areas, enabling machines to make more sense of data about those topics. For example, an ontology on the structure of an Army would enable a machine to understand that a Division is made up of Brigades, Battalions of companies and so on.

Other techniques that have been labelled as AI by some include Agent-Based Modelling and Genetic Algorithms.

Agent-Based Modelling is a method that creates "Agents" – discrete bits of code - that interact with other Agents through a set of rules. This approach is often used in building computer simulations of complex areas.

Genetic Algorithms is an approach that seeks to solve problems through an evolutionary approach that simulates the process of natural selection.



The Current High

AI has been through several highs and lows and is currently undergoing a high. This high is due to the application of machine learning techniques which we discuss later in the Biscuit Book.

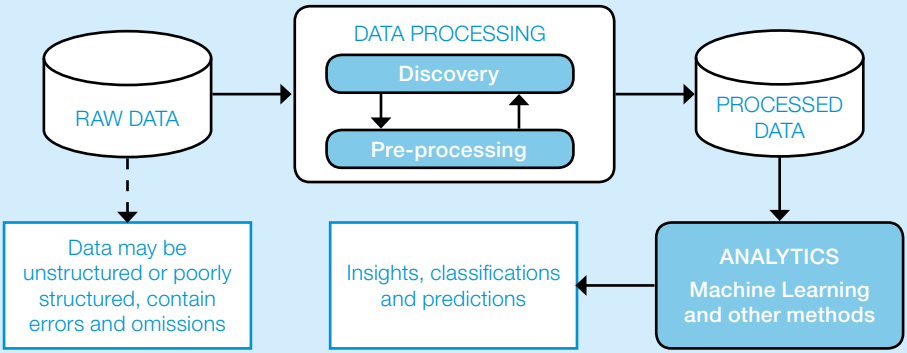
Data Science

Once upon a time this section might have been titled Big Data or Data Mining, but times and names move on. None of these terms are entirely equivalent but Data Science now effectively covers them all.

Stripped down to the core, Data Science is about the use of statistical methods encased in a blanket of data preparation and visualisation techniques. Exactly what is, and isn't, Data Science is of course somewhat debatable and about as well-defined as AI. Similarly, a Data Scientist can be anything between an expert in statistical methods, machine learning and visualisation techniques at one extreme, to anyone with a degree in maths or computer science looking for a better paid job (changing your job title to Data Scientist significantly increases employability).

Perhaps the best way to think about Data Science is as a process that takes data and generates insight and prediction from it.

A simple picture of this process is shown below:



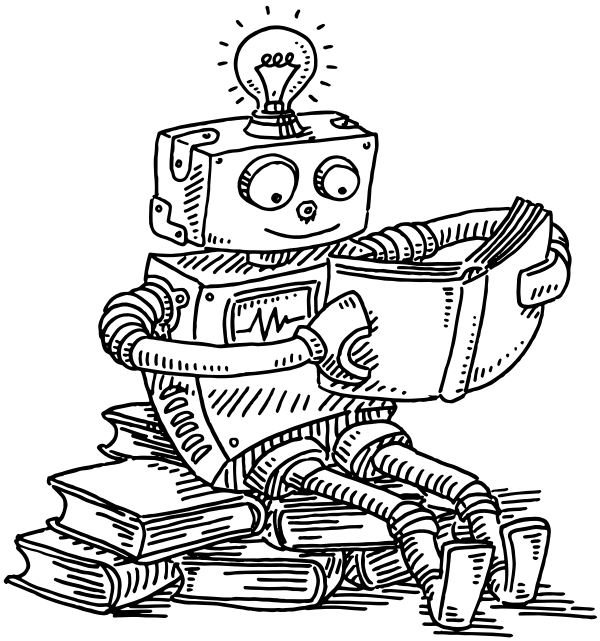
It should be stressed that this diagram is a simplification and there are many variations of what the process is, but this highlights the main elements. It should also be noted that Machine Learning is not the only way to perform analytics and that Machine Learning can exist outside of Data Science too.

Data

The one thing that can be said about most data is that it's hard. If you are in a situation where the data is easy you are either lucky or mistaken.

Some common reasons data can be hard:

- Difficult to actually get hold of it for commercial, legal, or other reasons, sometimes less rational – "it's mine you can't have it"
- Poorly documented
- Poor quality (very common)
- Not enough of what you need (even if there are large volumes of data)
- Ethically difficult to use (for example, it contains personal information)
- Difficult to combine with other data



Data Types

Data comes in different flavours, which may be one reason why some people are luckier than others. The main ones are: Structured Data, Textual Data, Digital Signal Data, and Image Data.

Structured Data:

Most data held in most databases are structured. Typically they are stored and represented as a table or more likely a series of inter-related tables. Common formats when output from a database include: Comma separated values (.csv), Excel files (.xlsx) and XML files (.xml, particularly useful if you have hierarchies in your data). One of the problems with much of this sort of data is that its quality may be quite poor, especially if people have been allowed to enter it. (As a general rule of thumb people who enter data into databases are usually more creative than those that design the database in the first place – generally this is not good for data quality.)

Day	Outlook	Temp.	Humidity	Wind	Play tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Text Data:

There is a lot of information held in text (that is after all what it's for) and such information is becoming more and more accessible to computers through Natural Language Processing (NLP).

Image Data:

This includes images and videos. In one sense it is highly structured: a picture being represented by a grid of pixels; but the image itself is highly unstructured and is difficult for machines to interpret.

Digital Signal Data:

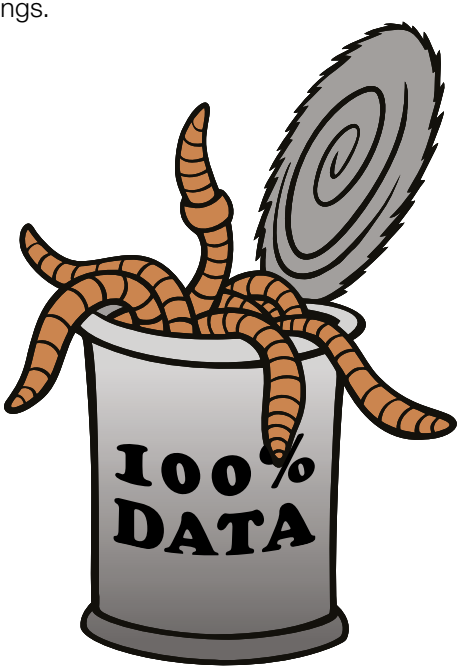
A Digital Signal Data is something in which time and amplitude have discrete values. It is obtained by sampling and quantifying a physical signal. Speech data is a special case of digital signal processing, in which the sound recorded is of one or more humans speaking.

Data Discovery

Analysis and Visualisation

Once you are lucky enough to actually get hold of some data, it will be necessary to understand what you have as it may not be exactly what it said on the tin.

Exactly how you undertake this voyage of discovery will be very dependent on the nature of the data. This could involve statistical measures, quality assessment, and perhaps visualisation techniques such as boxplots, histograms and scatter plots. For some data it may be important to understand the nuances of data, including understanding how it was collected and the way this may bias the data. For example, image data will be biased by the nature of the sensor used in the imaging, near infra-red sensors are very good at highlighting vegetation, so these will be detected with ease whist other objects may be less prominent. Such nuances may also include what the data actually represents – it may not be quite what you think, just because the data is recording vehicles, the collector's definition of a vehicle may be different from yours. This whole process may lead to some degree of disappointment but will also indicate work you may have to do to improve things.



Data Wrangling

Having discovered just how far short your data is from your optimistic expectations, all is not lost – there are things that can be done to improve matters. Even if your data was of good quality, it may still be necessary to process it so that it is suitable for the analysis tools you have, and perhaps also combine it with other data – a process known as conflation. These pre-processing stages are often called Data Wrangling.

Although tools and techniques exist to help this process, it is one where each data set will need to be approached in its own way. It has often been described as a cottage industry, although artisan may be a better description – if nothing else it should give an indication of cost.

It is often said, but never attributed, that Wrangling can take up 80% of the time of a Data Science project. Whatever the precise percentage for most projects it's a lot. Some may be tempted to cut corners because this process is expensive, time consuming, and gets in the way of the fun bit which is to do the actual analysis.

Our advice is to take this stage seriously, it's the only way to get good results; we'd only advise shortcuts in this stage for projects you don't want to succeed. This stage should also not be treated independently from the Discovery phase, in many cases they can be closely coupled – discovery highlights issues, fixing the issues discovers different issues and so on.

Data Wrangling Methods

Data cleaning: Data cleaning deals with duplicated or missing data, outliers, or noisy data (that is, data containing random values – think of the static you can hear on a radio). Duplicated data, where the same information occurs twice or more times, will need to be removed. A bigger problem is missing data. If too much data related to some item or measurement is missing it may be best to delete it. If not much is missing it may be possible to infer the missing values or use means or medians – these are not perfect solutions and may affect the results in some way, so care needs to be taken.

Outliers are instances in your dataset that are far from all the others. They normally occur when there is an error in the measurements. However, they can also indicate that you have a very skewed dataset (biased in some way). If the outlier is an error, you can delete the instance, however if it is a sample belonging to skewed data, deleting it will mean ignoring part of your data. Some data such as signal data may contain noise and it is often possible to apply noise reduction techniques. Other sorts of data may have inconsistencies which again can be systematically removed. An example is a dataset which may refer to the company IBM as “IBM”, “I.B.M.” or “International Business Machines”.

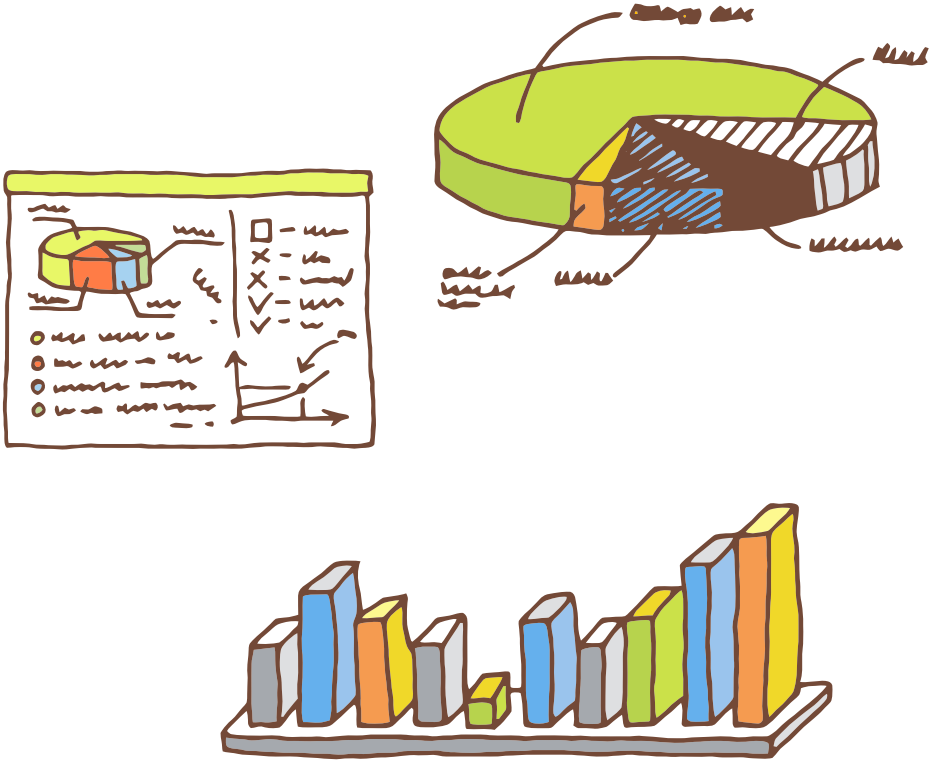
Data transformation and Data reduction: These are methods that you may want to apply to statistical data and are used to either transform the data to ensure all the data are in a comparable form (this is known as normalising) or removing redundant or unnecessary information (reduction).

Conflation: As if the process is not complicated enough already it is often necessary to merge components of two or more different datasets – this process is known as conflation. To add interest, an element present in two datasets may have two very different identifiers applied to it making it necessary to correlate the two using attribute data. Such processes may introduce errors though incorrect correlation and missed correlation, so care must be taken.

Analytics

Having obtained, discovered, visualised and wrangled your data, the time has finally come to perform analytics. Currently this is almost the same as saying the time has come for some machine learning, but this is not the only method of analytics that can be performed.

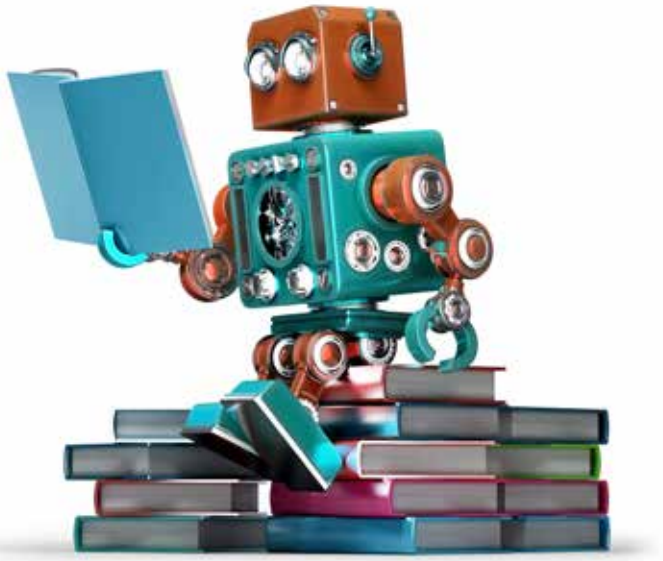
Before covering machine learning in the next section it's worth reminding ourselves that analytics were performed before the onset of machine learning. Classic statistical analysis is still alive and well, and specialised data such as geographic information is still healthily analysed using Geographic Information Systems (GIS). Such methods have advantages over machine learning especially when the amount of data is limited, and so are still worth considering.



Machine Learning

The aim of machine learning is to create systems (i.e. algorithms) able to automatically learn the relationship between input data and the classifications or actions you want to happen without being explicitly programmed.

Popular machine learning problems include object classification, tracking, image segmentation, audio recognition, or land-cover classification and there are many new and exciting problems where machine learning is starting to be explored, such as automatic subtitle generation, text generation, image colourisation and natural disaster prediction. You encounter machine learning when you speak to Alexa or unlock your smartphone using your fingerprint.



Machine learning differs from classical Symbolic AI because the latter requires a human programmed solution.

To help understand the difference, consider two robot mice trying to navigate a maze. The first has been programmed using symbolic AI and the second mouse has machine learning.

The first mouse has a lot of rules that tell it how to explore a maze to find a reward; someone has had to work out these rules. It will do quite well but if it encounters anything that the programmer didn't think of, like a new type of obstacle, it will get stuck.

The second mouse has to learn for itself and for quite a few runs of the maze it will be completely lost. But given enough runs it will work out the rules for itself although they will not be represented in the mouse's robot brain as we might imagine rules to be.

At present there are four main approaches in this field:

- Unsupervised Learning;
- Supervised Learning;
- Semi-supervised Learning;
- Reinforcement Learning.

Unsupervised learning is defined as learning solutions (known as models) that are not overseen (or supervised) by data labelled or tagged by someone - that is the machine has not been told what it is to learn. We give more detail on labelling later. The most popular family of unsupervised learners are clustering algorithms. Clustering is the process of finding groups in your data that are similar to each other is some way. Different clustering methods will have different groups defined by different sets of similar properties. An example is Google's unsupervised Image Classifier that learnt how to distinguish between cats and dogs. Unsupervised learning enables us to discover relationships that may be hidden in our data, for example, showing areas where particular activity one day may indicate particular crimes committed on the following day – something that may not have been obvious.

Supervised Learning refers to the family of approaches in which the learning is overseen by the label examples or the output of the data. The machine is given examples or training data (labelled data) of what we might be interested in, and also examples that are not what we are interested in. The model is then able to learn a function that maps the input of the training data to some output. So say that we want to teach a machine to identify cats in images. We do so by first labelling images of cats in as many images as we reasonably can. These images, along with a lot of other images that the machine is told do not contain cats, are then used to train the machine. Supervised methods obtain the most accurate results, since they explicitly tell the machine what is, and isn't, our thing of interest during the learning process. However, they also need the acquisition of labelled data for your dataset, which can be a problem if you require experts to label them or special equipment. In comparison, obtaining unlabelled data for unsupervised learning is relatively easy.

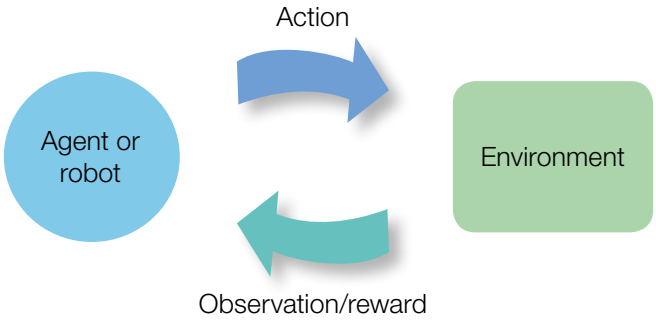
CAT

CAT

NOT CAT

Reinforcement Learning

Another alternative is to reframe the problem: in reinforcement learning, instead of training a model to find a function to link your input data to your label, you will be training an agent, which will make smaller decisions. For each decision, it will receive a reward. The goal then becomes to maximise the cumulative reward for the agent. Reinforcement learning has been used successfully in playing games where the game provides the "environment" and the high score (or some other way of telling how well you are doing) the "reward".



Classification and Regression

Depending on what you are trying to do, you will be dealing with either a classification, or a prediction, or regression problem. In classification, the goal is to identify a distinct class (the type of thing you are interested in, such as cars or cats). Popular classification problems include: object recognition, land-use classification, and image segmentation. On the other hand, if you are trying to predict a continuous value such as temperature, age, risk levels, etc., you will be dealing with regression methods.

Regression

Regression, or more properly regression analysis is a set of statistical processes for estimating the relationships among variables. Regression analysis helps us understand how the value of one variable changes when any one of the other variables is changed, while the other variables are held fixed. So in a simple example where there are just two variables X and Y, we might want to understand how Y varies as we vary X. So for example, you could work out how quickly your dinner will cook (Y) for different oven temperatures (X). Where there are three or more variables we still only vary one to see how another varies keeping all the others at fixed values.

Data and Machine Learning

The learning process typically has stages: training, validating and testing. Each stage will serve a different purpose and use different subsets of your data.

1. **Training phase:** In this phase you will use a subset of your dataset, referred to as the training set, to learn a model. Think of the training set as the set of examples that you give your machine learning algorithm to learn from. In supervised learning, this data will need to be labelled so the machine can learn from the labels. Labelling can take a considerable amount of time and if it is poorly labelled the results will also be poor so it is important to get this right.
2. **Validating phase:** In order to gauge how well the training has worked you can expose the model to further data which will optimise the results. One possible problem with training is that the data can be "over-fitted" (become too sensitive to certain aspects of the data) which can lead to poor results. For example if we want to identify images that contain tanks then it is possible that the training data will lead the model to be sensitive to green things in general. The validation data set can help to correct for this. It can also help to indicate when enough data has been used and no more training is necessary.
3. **Testing phase:** In this phase, the model is tested by giving it the test set, a set of unseen samples and seeing how the prediction of your model compares to the actual output you were expecting. The more similar these are, the better your model is at "fitting your data".

One important thing to remember is that it is crucial that you ensure that no instance in your dataset is used at more than one stage - to do so would invalidate the testing phase, a bit like getting the same question in a real exam as you did in the mock. The other thing to appreciate is that this can require a lot of data, usually the more data you have the better although you will get to a point where the returns begin to rapidly diminish. If you don't have a lot of data then there are some techniques that can be used, usually referred to as "lowshot learning" techniques, but it may be that machine-learning isn't the best approach.

Machine Learning Methods

To give you some idea of the sheer number of methods that are employed in Machine Learning here is an incomplete list:

Auto Encoder (AE)
Boltzman Machine (BM)
Convolutional Neural Net (CNN)
Decision Trees (DT)
Deconvolutional Network (DN)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Deep Feed Forward (DFF)
Deep Q-networks
Deep Residual Network (DRN)
Denoising Auto Encoder (DAE)
Echo State Network (ESN)
Extreme Learning Machine (ELM)
Feed Forward (FF)
Gated Recurrent Unit (GRU)
Generative Adversarial Network (GAN)
Hopfield Network (HN)
Kohonen Network (KN)

Liquid State Machine (LSM)

Long/Short Term Memory (LSTM)

Markov Chain (MC)

Neural Turing Machine (NTM)

Perceptron

Radial Basis Network (RBF)

Random Forest (RF)

Recurrent Neural Network (RNN)

Restricted Boltzmann Machine (RBM)

Sparse Auto Encoder (SAE)

Support Vector Machine (SVM)

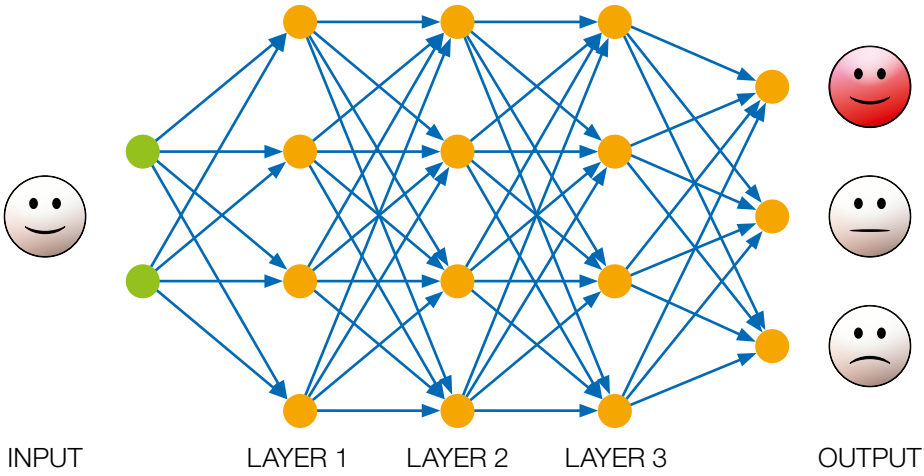
Variational Auto Encoder (VAE)

… and so on.

And, of course, this list is growing all the time. These are just the methods. The number of tools, whether commercial or open source, that use them is much greater.

Artificial Neural Networks

Many of the listed methods fall into a class of methods known as Artificial Neural Networks (ANN) as they are based on the principles under which our own brains are wired. Artificial Neural Networks are the most common form for Machine Learning methods nowadays and the ones where most progress has been made.



Structurally an Artificial Neural Network comprises a number of layers of nodes (the neurons) each connected to nodes in the next layer. The first layer is known as the input layer, the last is the output layer and the intermediate layers are hidden layers. Crudely, the way the neural networks work is that each connection between neurons is weighted and these weights are adjusted until the desired output matches that of the input. Each output neuron represents a probability as to how well the output matches the input. In the diagram above the ANN will be trained by showing it a series of happy, sad and neutral faces with no two faces the same. Each face will adjust the weights in some way until we get to a situation where presenting a happy face will result in the output registering it as happy.

Deep Learning

Deep learning methods are a family of Artificial Neural Networks which have been consistently obtaining state-of-the-art results in most machine learning tasks since their development. Although deep learning networks can be used in unsupervised and reinforcement learning, they are generally used more frequently in supervised learning problems.

They work by using a succession of multiple layers where each set of layers effectively acts as a neural network its own right. So for example if we have an image processing problem, the first set of layers might take as input the raw pixel data and output basic shapes such as lines of various shapes. The next layer may take these and output specific shapes such as circles and rectangles and so on until the output comprises different types of vehicle.

Depending on your type of input data and the type of problem you want to solve, you might find some types of deep networks better than others.

Tabular Datasets:

Not that many deep networks can be used for tabular datasets because most popular networks use spatial information (i.e. which attributes are next to each other) in order to extract features automatically. If you encounter tabular data, basic neural networks will be your best initial dataset.

Audio/Video/Temporal datasets:

If you have temporal information (i.e. data that varies through time) recurrent neural networks (RNN), and long-short-term memory networks (LSTM) are two of the best candidates.

Image Datasets:

Most of the advances in deep learning have been done for image datasets. If you have images that you want to classify, segment or track, good deep learning candidates include convolutional neural networks (CNNs).

Appreciating the Complexity

In order to appreciate just how complex things are getting, here are explanations for two popular methods in more detail than we have used elsewhere. The point is not to understand the method as such - that is a bonus. The point is to understand just how complex and specialised the detail actually is.

Before the detail, the long story short:

Convolutional Neural Networks are Artificial Neural Networks on steroids.

Long Short-term Memory Networks are Artificial Neural Networks on steroids with a time machine.

Convolutional Neural Networks (CNN)

Mostly applied to visual datasets and in natural language processing, CNNs are one of the most popular deep networks. CNNs consist of an input layer, multiple hidden layers, and an output layer.

The input layer is generally the image (or images) in your training set. The hidden layers mostly are convolutional layers, rectified linear unit (ReLu) layers, pooling layers, and fully connected layers (gasp!).

Convolutional layers, the core block of a CNN, learn filters that activate when specific features are recognised somewhere in the image. Pooling layers are sub-sampling layers that partition the input image into non-overlapping windows and output the maximum of each region. ReLu layers are used to remove negative values from an activation map by setting them to zero. Fully connected layers are where the classification is actually done. They learn which activations relate to which classes.

Finally, you will have an output layer which will return your prediction, and a loss layer, which specifies how training penalises the deviation between the prediction and the actual true output. Depending on your problem, you can choose different loss layers. Softmax loss, for example, is used for predicting a single class out of N mutually exclusive classes, while Euclidean loss is used for predicting (regressing) continuous values.

CNNs are the base of many other methods, such as Deep Q-networks (used in Reinforcement Learning), Fast RCNN, Fully convolutional neural networks, amongst others.

We did say it was complicated!

Long Short-Term Memory Networks (LSTM)

LSTMs are a variation of Recurrent Neural Networks (RNNs), a type of deep network suited, not only for single data points (such as images or tabular data), but also sequences of data (such as video or speech). While they were introduced in 1997, making them relatively old in Machine Learning terms, they still maintain state-of-the-art results in very popular machine learning problems involving human action recognition and speech recognition. They have also been used in some unusual problems such as automatic music composition and caption generation for film and TV.

LSTMs are composed of input gates, output gates, cells and forget gates. A cell is able to remember past values at different time intervals, and the forget gates control the flow of information in and out of the cell (i.e. what needs to be remembered and when it needs to be applied).

Low-shot Learning

Due to the need of networks to first learn the features of your data, deep learning networks have traditionally needed vast quantities of data. Current state of the art results in image recognition and audio classification use millions of labelled data points, which is not always achievable.

As a response to these impossible requirements with regards to getting data, a new area on deep learning, called low or few-shot learning, has been gaining attention in the last few years. Low-shot learning aims to automatically learn features from datasets in which each class has very few samples, generally twenty, ten, five or even one. Results are still not able to improve those from networks trained in millions of samples, but they are still very competitive.

Measuring Success

We've got the data, pushed it through a Machine Learning model and now we've some results. Thing is, are they any good? There are several types of metrics that can be applied to measure the level of success depending on the problem that you are working on. The most commonly reported are Root Mean Square Error, the confusion matrix, recall and precision.

Root Mean Square Error (RMSE): This error metric measures the differences between the values predicted by your model and the values observed in your dataset. It is defined as the standard deviation of the prediction's errors (often referred to in the literature as residuals). Due to its nature, RMSE is particularly good for prediction problems.

Confusion matrix: If you are dealing with a classification problem one way to gauge the performance of the model is through a confusion matrix. It allows you to see the "confusion" between the classes, that is which class is commonly mislabelled as another class. (We hope that's not too confusing!) If your problem is binary (i.e. you have two classes), then your confusion matrix will look like the one below. Your confusion matrix will have the following information:

- True Positive (TP) : Observation is positive, and is predicted to be positive.
- False Negative (FN) : Observation is positive, but is predicted negative.
- True Negative (TN) : Observation is negative, and is predicted to be negative.
- False Positive (FP) : Observation is negative, but is predicted positive.

Confusion matrix for binary problem

		Actual class	
		Class 1	Class 2
Predicted as	Class 1 e.g. male	TP	FP
	Class 2 e.g. female	FN	TN

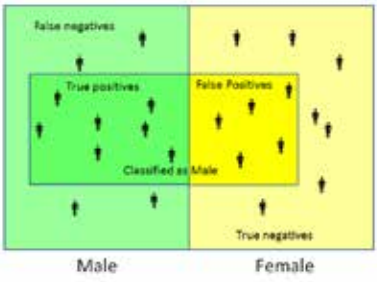
The above matrix shows just two classes. The more classes the more complex the confusion matrix gets. We won't show more complicated examples lest it becomes too confusing.

Accuracy: It is the rate of correctly classified instances or, in other words, how many times your classifiers were successful. It can be calculated as:

$$\textbf{Accuracy} = (TP+TN)/(TP+TN+FP+FN)$$

Precision: Represents the total number of selected items that are relevant. So it is the true positives divided by the all those selected as positive (true positives plus the false positives).

Recall: The number of correctly classified elements of one class (true positives) divided by the total number of elements labelled as belonging to that class (true positives and false negatives). It is more restrictive than precision, since false positives (i.e. misclassifying something from another class as that class) are penalised.



So in the example above where a machine is trying to classify males (perhaps from photographs) the precision is:

$$\frac{\textbf{7 Males (TP)}}{\textbf{7 Males (TP) + 5 Females (FP)}} = \textbf{0.583}$$

And recall is:

$$\frac{\textbf{7 Males (TP)}}{\textbf{7 Males (TP) + 4 Males (FN)}} = \textbf{0.636}$$

In both cases, the higher the value the better so these results are not terribly good, we wouldn't advise buying this solution.

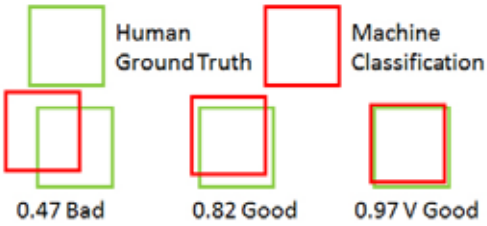
F1: Is useful to see the relationship between precision and recall at the same time. It is defined as:

$$\textbf{F1 = 2 x (precision x recall)/(precision + recall)}$$

Again, the higher the value, the better, 1 is perfect (and probably hard to believe)!

Jaccard Index or Similarity Coefficient

The Jaccard Index is a statistic used to measure the similarity between different sample sets. It is defined as that part of an image that is common to both the ground truth and the machine classification divided by that part of the image that is in one or the other. The larger the value, which will be between 0 and 1, the better.



The Jaccard Distance, which is the complementary of the Jaccard Index, measures the dissimilarity between the two areas so the higher the score the greater the difference.

Limitations

AI, Data Science and Machine Learning are great but not perfect. That there are limitations might not be clear, and with all the hype it is easy to misunderstand the capability and over-estimate what is possible. Where they have been successfully applied, it is often in a more limited sense than may be apparent from the way it is reported. So when reading reports on how these areas have aided response to humanitarian disasters, prevented crime, diagnosed illness and a host of other things, it is necessary to understand that the achievements are often very specific to a particular set of circumstances.

We know from a personal perspective that things do not always work as we'd like: how often has Siri, Alexa, Google misunderstood you? How often do Amazon and Netflix recommend something that does not match your taste at all? Have you ended up in an unexpected traffic jam after following your SatNav? Have you ever had to end up calling your bank's customer service because the chatbot was not answering your questions?

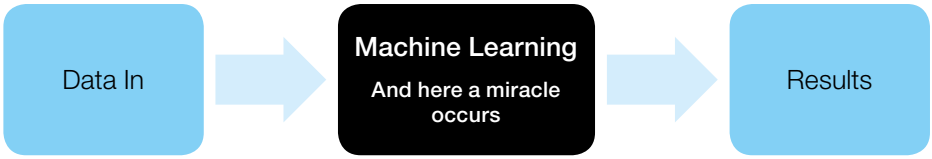
Thus an appreciation of limitations is as important as understanding strengths.

This is because even the most advanced methods have one clear limitation: they are all data dependent. Most Machine Learning models and Data Science methods are trained on large, annotated data sets. These annotations, more often than not, are generated by people. The quality, completeness and correctness of the data, and the annotations will directly affect the quality of your results. Think about it this way, when you do not know how to do something, you look for examples on how to do it. A Machine Learning model will require the same: if it does not have enough examples of what classes you want to classify or predict, it is impossible for it to learn successfully to do so.

Examples of the harm of unbalanced, biased or incomplete data are as numerous as concerning: in 2014, a company created an ML-based system to automatically filter CVs to optimise hiring procedures and eliminate biases. However, since the system was trained on data from the company over a 10-year period, and hiring practices during that period were biased, the resulting system further supported said practices. Investigations revealed that the system penalised CVs with words like "women's", "women's chess club captain", and rejected graduates of two all-women's colleges.

The Big Problem with ML

Put bluntly, the big problem with machine learning is that nobody knows what it's doing!



The machine learning phase is very much a black box, we understand the general principles of Artificial Neural Networks but we don't understand what is actually going on in any detail when it is being applied – for all intents and purposes it's magic!

You may ask why this is a problem, after all its producing good results so why worry? The problem is that because we don't know what is going on, we can't help fix things when they go wrong and it is difficult to understand the answers we're given. This has serious implications when it comes to where we can apply Machine Learning. If all we're interested in doing is scanning the web to find pictures of kittens, this is not really a problem. But, if we want to apply machine learning to control an aircraft, then we really do need to understand what is going on. This is one area where symbolic AI and conventional programming has a massive advantage because, in these situations, we do know how they do what they do. Such approaches are thus transparent. There is a lot of work currently underway in the area of transparency and machine learning, this is known as "explainable AI", but we are still far from a solution. At best we are able to determine which pixels the machine used to classify a picture of a dog as a dog – but its choice of pixels will be a bit weird to us and certainly not very informative as to how it is working.

Wrapping Up

We hope your tea hasn't gone cold, that you enjoyed your biscuit and, most of all, we hope you found this a useful and fun way of learning about AI! AI has been called the "new electricity" and is a technology that we believe is going to power an increasing number of our industries and impact many aspects of our day to day lives. Combining an understanding of AI with expertise about specific application domains and a strong sense of ethics is the key to knowing how to use these powerful techniques in a manner that creates value and makes our world a better place for everyone.

Working with Dstl on AI

Dstl works with its partners and suppliers to deliver a lot of research on AI, Machine Learning and Data Science. This research aims to understand how we can responsibly & safely apply these techniques to a wide range of Defence and Security challenges including military decision making, autonomous platforms, computer network defence, sensing, defence logistics, policing and security, streamlining back-office functions and a whole lot more.

To make it simpler to engage with us on AI, Dstl has established AI Lab, a pan-Dstl flagship for AI, Machine Learning and Data Science. AI Lab works across Dstl's entire portfolio of research and aims to establish a world-class capability in the application of AI related technologies to Defence and Security challenges. If you'd like to get in touch with AI Lab you can email:

ai_lab@dstl.gov.uk

Content includes material subject to © Crown copyright (2019), Dstl. This material is licensed under the terms of the Open Government Licence (OGL) except where otherwise stated. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gov.uk.

All third party images reproduced in accordance with their associated Copyright Licence Agreement. The terms of the OGL do not apply to any incorporated third party content.

DSTL/PUB115968

Ministry
of Defence