Introduction to artificial intelligence and machine learning

Introduction

Artificial intelligence (AI) and machine learning are related emerging technologies. The terms are often used interchangeably. Machine learning is a current application of AI based on the idea that machines should be able to learn based on their access to data.

Artificial Intelligence

AI is a loose term for computers using processes that are traditionally associated with the human brain. These range from basic functions such as pattern recognition and analysis, to higher-order cognitive phenomena like reasoning, evaluation and decision-making. AI covers a range of overlapping tools, approaches and functionality, that include:

- machine learning or deep learning - recognising patterns in datasets and creating predictive models that can be applied to new data
- neural networks - computer systems that approximate the physical structure and learning behaviour of neurons in brains
- robotics - hardware that is capable of carrying out physical tasks either automatically or under human control
- automation - software or hardware that is designed to carry out highly structured, repeatable tasks

Even the most sophisticated technologies are sometimes considered as ‘weak’ AI because they are specialised to carry out a certain task or range of tasks. It’s important to note that a ‘weak’ AI may still be better than humans at performing a specific task. This is in contrast to ‘strong’ AI or artificial general intelligence (AGI), which would be functionally similar to a human and able to learn and perform any task a human could.

Machine learning

Machine learning grew out of early work on artificial intelligence. It still forms an important part of that wider discipline but is also an increasingly prominent data science tool in its own right.

Machine learning is when a computer recognises patterns using statistical models and improves its predictions based on sets of historical data. This is often done iteratively using training and testing datasets. The series of mathematical or logical steps used in this process is called an algorithm.

These processes are now commonplace in our daily lives and are most applicable in predictive analytics, a subset of data science. Examples are the user recommendations on internet shopping, music and video streaming websites.

Supervised and unsupervised learning

You can broadly separate machine learning into 2 categories: supervised and unsupervised. In supervised learning, data is labelled to allow the programme to analyse how inputs are related to outputs. This is particularly useful when an organisation has some knowledge of cause and effect but needs to improve the accuracy of this understanding. This is achieved by:

- choosing a data set that is representative of the real-world scenario and separating it into two parts
- using the first part of the data to create a model that answers a question or set of questions (a process known as training)
- evaluating the accuracy of the model by checking its answers when it is run against the second part of the data
- feeding this evaluation back into the model so it can continue to improve its accuracy when confronted with real-world data

This feedback loop enables the model to learn from experience, rather than explicit programming. For example, email providers know that messages which fulfil certain conditions (such as being sent from a particular email address) are likely to be spam. Supervised machine learning can be used to refine this categorisation and improve inbox filtering.

Unsupervised learning does not involve the specification of expected outputs. Instead, programmes are designed to detect patterns in data, even though the designer may not know in advance what these patterns could be. There are many varieties of unsupervised learning, but a common example is finding trends in meteorological data that can help in weather-forecasting.

Augmenting human decision making

Machine learning programmes offer greater flexibility than conventional automation (such as macros in Excel). Traditional automation relies on using a scripting language to rapidly execute predictable, repeatable tasks on consistently structured data. But machine learning programmes can be trained to process data even if it is relatively unstructured or in unpredictable configurations. This means they can be used to analyse large sets of diversified data much faster than people, and so aid human analysis.

Machine learning is not a perfect predictor. Much like humans it can make mistakes based on its interpretation of the data. At present, it can sometimes replace a human decision-maker. However, it is often most valuable as an added step in the human decision

making process that provides useful models and analysis of data for the user. For example, pattern recognition can be used to flag records that might need further investigation, rather than providing a definitive decision on its status.

Things to consider

Organisations should:;

- clarify the user need for AI or machine learning - sometimes traditional approaches may be more effective
- consider any public concerns about the technology, and implications for their workforce
- think about bias created through:
    - availability/accuracy/selection of the data used to train the model, and whether the model will be as accurate with real world data.
    - type of model used and the transparency of their internal operation
- understand that the outputs of machine learning are only as good as the data input
- make data scientists more visible within organisations - this will improve statistical literacy and understanding around the technology
- use the principles set in the Government Data Science Ethical Framework
- consider the GDPR and their obligations to the public on the use of data
- balance the accuracy of the algorithms with wider fairness and policy objectives