ps ai labs

Artificial Intelligence (AL) / Machine Learning (ML) / Data Science
A Primer for Executives

Introduction

This is a short primer (10 min read) for executives evaluating how best to apply AI into their business. Given the scale and fast-moving nature of this field, we recommend reaching out to discuss these ideas in a collaborative value workshop, but we hope this briefing paper gives at least some ideas on key questions to ask.

AI Tools and Approaches

Everyone is interested in AI. Perhaps it's because, according to Microsoft, companies that exploit their data through advanced applications of AI and ML are outperforming the others by 11.5%. But many business executives who could benefit from AI don't really know what it is. So before jumping into where best to use it, let's define some terms:

Artificial Intelligence (AI) is a broad term coined in 1956 in the now famous Dartmouth Summer Research Project on Artificial Intelligence. Today the term is now used to encompass a wide variety of computer techniques and technologies that attempt to address aspects of "intelligence". These include abstract thinking, vision, understanding natural language, reasoning, problem solving and decision making.

Currently many commentators divide AI into *narrow* AI, that is to say techniques that solve for a specific and therefore narrow problem (for example spam email filters, grammar correctors or recommendation engines that suggest which film to watch next) and *broad* AI which attempts to solve for more complex problems such as driving a car.

Machine learning (ML) is one technique that allows a computer to learn from data without programming it with hard coded rules. In essence the machine *learns* from the data itself rather than being "told" what to do.

This process of learning falls into two types. *Supervised learning*, in which case the prediction is known and is based on looking at other independent variables. For example: How much is my car worth today? This might be based on its age, condition and what I bought it for. In other cases the machine is carrying out *unsupervised learning,* in which case you are not trying to predict a known variable such as the price of a car, but rather find what patterns already exist within the data that help you identify groups or clusters within that data. For example: what patterns can you find from 10,000 cars? One might be those that are most likely to break down. The data set may already have information such as last service date and the age of the car which can act as a predictor of breakdowns.

It is called *unsupervised* because there is no predefined structure to the outcome. As an analogy, painting by numbers could be called *supervised* and some blank paper and coloured pens *unsupervised*.

Machine learning typically looks at two types of *prediction problems*. The first is *regression problems*, that is to say predicting a numerical output. For example: What will the price of a car be tomorrow? Secondly *classification problems*, where you have to predict a category (rather than a number). For example: Will the next car passing be a Ford or a Nissan? The mathematical prediction in both cases is based on a set of input data/variables rather than a set of hard coded rules. So, the "car price prediction" could be based on the last 5 years' worth of car prices, and "which car manufacturer" might be a record of all the models and manufactures of cars that passed by in the last week.

To process the data and all its variables, mathematical *models* are created. A model is sometimes also called an *algorithm*. Essentially it is a set of steps, with a beginning and an end, used to predict an outcome based on past behaviour. Machine learning uses many model approaches. The most common are:
1) Linear, where the calculation tries to find the best fit for the data along a line.
2) Tree based, where possible outcomes are laid out in the form of a tree and its branches.
3) Neural/deep learning, where more complex connections are laid out in layers, and the model learns about the relationship between the inputs to outputs.

Each of these model approaches has advantages and disadvantages depending on the data you have and prediction problem you are solving for. These models typically calculate a confidence level in how accurate they think the prediction will be. Since this is almost always <100% (and for various excellent ethical and pragmatic reasons), the concept of augmenting human intelligence with these models, rather than replacing humans with them, is a sound one.

Data science is a broad term to encompass the skill sets and approaches required to design, write, evaluate, and run machine learning and other AI techniques. Dr. John Elder, one of the founding fathers of data science, noted that data scientists are obsessed with one question above all else: "How likely are we to have got this result just by chance?" In the enterprise setting there are a number of closely aligned disciplines that overlap with data science -- data strategy, data engineering, and data analytics and visualisation.

Applying AI and Machine Learning

There is a virtually unlimited range of use cases in which AI/ML techniques can be applied in your business. So where to start? Here are three questions you can use to help you identify use cases that make the most business sense to pursue.

1) **Which business problem?**

The research from McKinsey Global Institute: "Notes From the AI Frontier - Insights From Hundreds Of Use Cases", showed **two-thirds** of the opportunities to use AI are in improving the performance of **existing analytics** use cases.
https://www.mckinsey.com/featured- insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning.

This was a study of 400 use cases across 19 industries and nine business functions. In 69 percent of the use cases they studied, deep neural networks can be used to improve performance beyond that provided by other analytic techniques. Cases in which only neural networks can be used, which they refer to as "greenfield" cases, constituted just 16% of the total. They estimate that modern deep learning AI techniques have the potential to provide a boost in value above and beyond traditional analytics techniques ranging from 30% to 128%, depending on industry.

So, the good news is that focusing on existing areas in which you are already using analytics is a legitimate approach.

With a mental list of such existing areas in mind, there are four ways to identify worthwhile business problems to apply AI/ML techniques to that I will mention.

The first is to look at the **"to be state".** Ask yourself:

- If I am able to predict this <xxx> better then so what?
- What decisions can I improve by having "better" data?

Secondly, you can take a more **process or workflow driven** approach. For example look at how people interact at work or in a particular process and ask:

- Which types of interaction could benefit from AI/ML approaches?

Interactions can fall into 6 types: i) Monitor (observing what is going on), ii) Manage (assigning work), iii) Execute (completing work), iv) Communicate (telling people what is going on), v) Analyse (critical assessment), vi) Decide (making a choice between options).

Thirdly, worthwhile business problems are strongly linked to business objectives and KPI's such as those shown in the table:

Data & AI / ML investments must drive Business Objectives & KPI's

Revenue	Efficiencies	Customer Exp
Visits & conversion	Staff productivity	CX / Engagement
Customer sales uplift	Staff reduction	Brand
New customer acquisition	Effort avoidance	Reputation
Sales frequency	Cost avoidance	Net promoter score
Price / margin optimization	Asset optimization	Competitive position
Lifetime value maximization	Efficiency	Insights

Lastly always make sure you ask: Can we actually action it? Do we control the levers? For example, there is not much point in calculating prices dynamically if they cannot be changed in real time.

2) **Is the data available?**

In simple terms you can only predict with any sort of confidence level if something is going to happen if there is a history of the thing happening in the past (this is the famous "unknown unknowns" referred to in the Johari window technique devised by **Jo**seph Luft and **Harri**ngton Ingham). Trying to predict in areas with no precedence is not recommended.

There is no standard rule for how much data is needed but data scientists generally advise the more the better. That means a distribution (over time) and richness (across dimensions and labelling of variables).

A label is the value of a variable e.g. "car colour" is the variable, the label is "blue".

Of course, the data should also be representative of the problem you are trying to solve for. If the data is skewed, your predictions will be skewed too. Data scientists will typically look for training data to generate an algorithm, input data to produce a prediction, and feedback data to continuously improve it.

In reality it is not always easy to define what is the "right" data, hence why data is a science and involves hypothesis and experimentation.

IDC estimates that ***"Only 0.5% of all data is analysed"***

The easier cases are where the outcome you want to predict is easily identified or labelled. For example: after looking on the website, did the customer buy the product? If you don't definitively know the outcome, you still may be able to derive value from the analysis or you may be able to find a proxy within the data to suggest the outcome happened. For example, if you don't know if the customer bought the car (this would be called unlabeled data) but they did book a test drive. Inconveniently sometimes all the data you need to solve for a problem might not exist within your business, and you may need to source it externally. This is called augmenting or enriching your data.

Some data science projects can burn 80% of effort in getting and cleaning data, so without a doubt putting time upfront into evaluating your data will be time well spent.

It has also been estimated that 80% of the world's data is unstructured, for example in images and video, free format text documents etc. Although this data may be more difficult to analyse, some of the most exciting value cases can lie when this is combined with structured data that sits in tables and databases. For example, collating unstructured medical records with ECG and structured data from other bio-medical devices.

3) What if it goes well (or not so well)?

In our experience, experimentation is the way to go. It has worked in nature for millions of years, so no reason to suggest it doesn't work for AI/ML.

"In an increasingly digital world, if you don't do large-scale experimentation, in the long term — and in many industries the short term, you're dead"
Mark Okerstrom, the CEO of Expedia Group

In practice that means being prepared for failure and creating a culture of experimentation where you are testing hypotheses rather than relying on the ways things are done today.

A mature experimentation mindset would mean you have:

- Business stakeholders embedded into the experimentation team
- Quick autonomous decisions
- Continuous customer feedback loop integrated into process
- Automated experiment scheduling
- Data cleansing / Outlier detection
- Sequential testing techniques

> "Our success is a function of how many experiments we do per year, per month, per day"
> - Jeff Bezos, Amazon CEO

However, the inconvenient truth is that even if your AI/ML experiment "works" it may not actually see the light of day. Venture Beat have estimated that only 13% of machine learning models make it into production.

Assuming the business case is "worth it" and the models work, you may still need to address these common hurdles:

- Unreliability / inexplicability of the model
- Regulation
- Existing processes or ways of working unable to effectively use new approach
- Lack of data
- Performance or scaling difficulty
- Unclear business case

Summary

There are plenty of use cases for AI/ML, many of which can be applied to your existing processes and problems. Take some time to assess the business value of, and your ability to take actions with, any AI/ML investments you make. Also, check you have access to appropriate data, and watch out for all the non-data elements that might slow you down such as legal, compliance, existing ways of working, etc. But if you can work through that, the size of the prize is definitely impressive.

Good luck with your AI/ML initiatives, and please reach out to info@psail.com if you would like to set up a value workshop or discuss anything in this document.