CHECKLIST

IS MACHINE LEARNING THE RIGHT APPROACH FOR YOUR PROBLEM?

1) CAN YOU DESCRIBE YOUR PROBLEM IN TERMS OF 'INPUTS' AND 'OUTPUTS'?

It's important to be able to formulate your "expected output" – the target outcome for the problem you're trying to address – and your "inputs" – the variables in your data that affect that outcome.

For instance, "with the available resources, can I achieve better performance"? Or, "what additional resources do I need to achieve better performance"?

In this example, performance is your output and resources are your input. To describe your problem like this, you need to understand what role your variables might play in the target outcome for the problem you're investigating. You might want to draw on expert knowledge in the field to help you do this.

2) CAN YOU FORMULATE AN OBJECTIVE FOR YOUR PROBLEM?

For instance, if you're a manager with available data on resources and productivity, your objective might be "I would like to increase productivity by 20% in 6 months."

It's essential that your objective is both measurable and achievable. Utopic targets can cause problems later down the line: they may lead to poor algorithm performance because unrealistic targets cannot be modelled or predicted with existing data sources.

3) IS IT POSSIBLE TO SOLVE YOUR PROBLEM WITH MACHINE LEARNING?

Once you've confirmed that your objective is achievable and measurable, now you need to determine whether it is achievable using machine learning.This step requires basic data analysis understanding of what the machine learning algorithms can achieve, so you'll need to draw on data expertise to help you.

In short, there are some standard types of problems that are suitable for a machine learning solution — and you need to work out whether your problem is one of these. For instance, finding patterns in data, grouping data or predicting the future from a sequence of past data. If your problem can fall into one of these categories, then machine learning could be an appropriate tool for solving it.

Objectives with a predictable aspect to them are generally suited to machine learning. In this case you'd start with the input data and some known results. For instance, "with x resources, productivity is y." Then you'd build a model that learns from these known results and can be reapplied for any unseen, similar input data.

However, machine learning is less appropriate for objectives that are iterative processes — a sequence of steps that are applied to get to the solution. Here you'd start with the data and then build a step-by-step algorithm to obtain the result.

4) IS IT NECESSARY TO USE MACHINE LEARNING TO SOLVE YOUR PROBLEM?

In some situations, statistical methods can be applied to achieve your objective and so it may not be necessarily to develop and apply a machine learning model.

It all depends on the nature of the problem at hand. Experts can help you understand what type of problem you're facing: is it a regression, classification, clustering, rule association or decision making problem? In some of these cases, there may be no need to apply a machine learning algorithm.

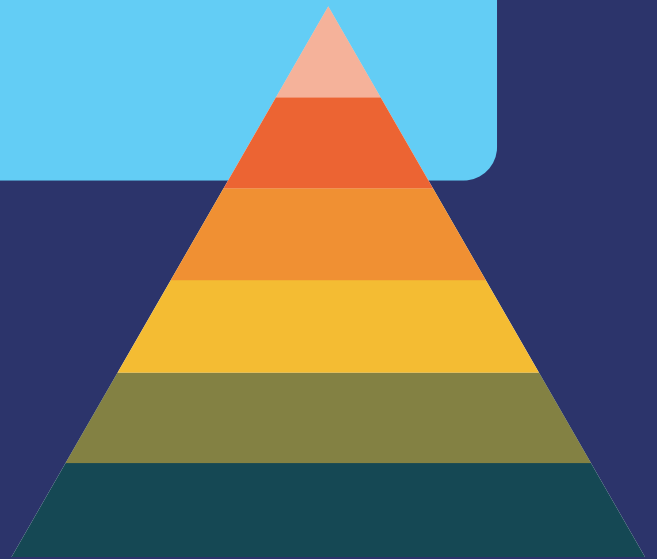
For instance, simple 2D or 3D plots of variables together with basic correlations might be enough to achieve your objective. But, if you want a template you can apply anytime you get new data added, or you want the model to continuously adapt to new data you aquire, then a machine learning model may be more accurate.

5) DO YOU HAVE AN ACHIEVABLE GOAL FOR YOUR PROJECT?

While the objective we set in question two should outline what you want to achieve with respect to your problem, your project goal should describe how you want to use machine learning to achieve it. You need a purpose for your project — the goal should articulate how you're going to use machine learning to address your problem.Do you want to improve an existing system or develop a completely new approach to your problem?

For example, perhaps improvements have previously been done "manually" by an expert in the field. They may have been changing some parameters in a bid to improve the output. For instance, in the productivity example, perhaps the expert has been changing the resources slightly, on a trial and error basis, to get better performance.

In this case, your goal might be to use your machine learning model to find a mathematical relationship between your inputs, or "variables," and your output. This could be expressed either as a rule — "if the resources are x, then the productivity level will be y" — or a mathematical formula. If you achieve this goal, you could then apply the formula to make predictions rather than relying on trial and error.



6) CAN YOU DEFINE A SUCCESS MEASURE?

This could be a binary success or failure metric, or it could be a percentage that reflects the degree to which you've achieved what you set out to.

These measures could be based on both the extent to which you achieved your project's goal and how well your machine learning system performed. How close did you get to the goal? And, how fast was your system? What resources did it require? How easily could you fine tune the parameters? How easy is it to interact with it?