

6 Steps to Jumpstart Machine Learning Using the Resources You Already Have

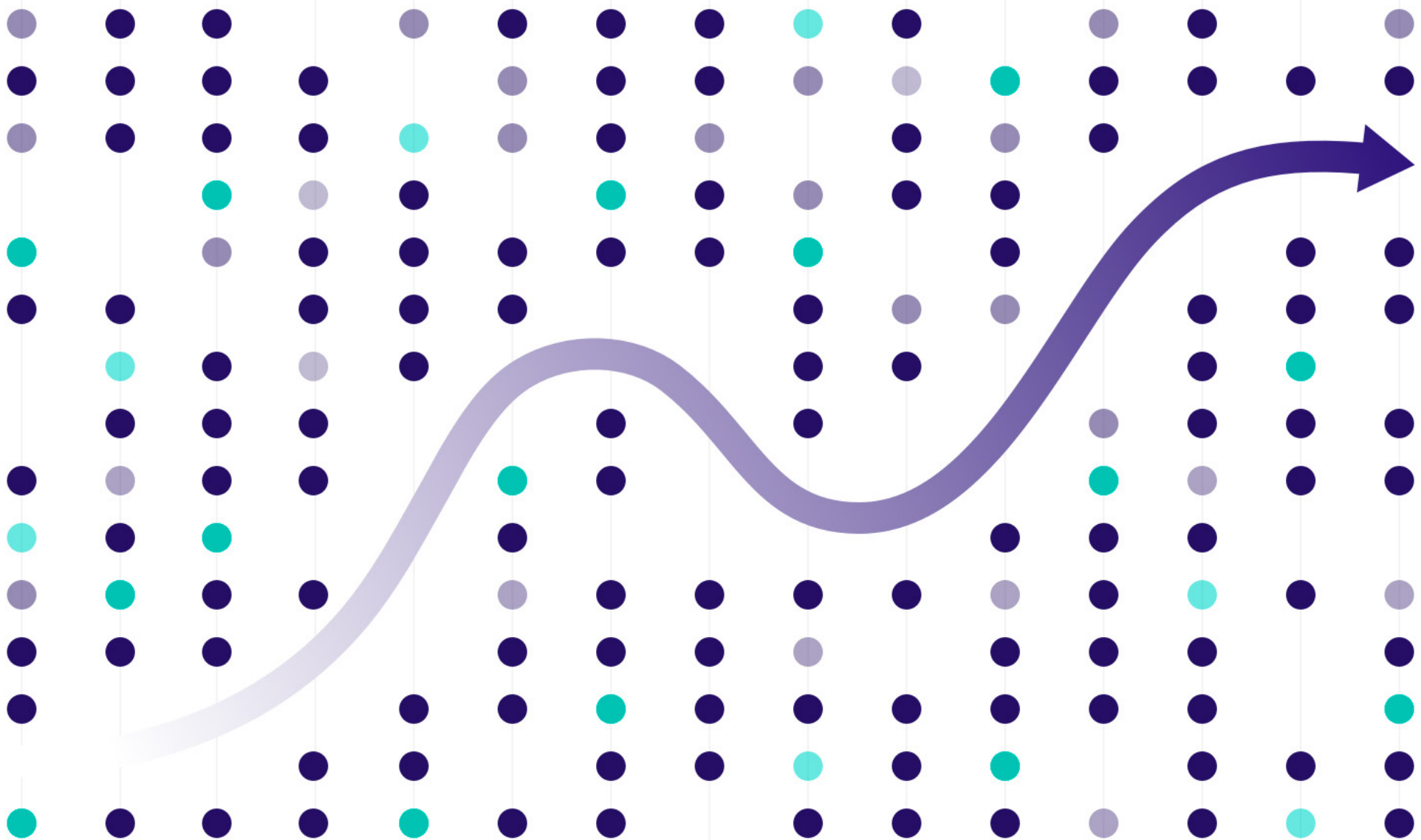


Table of contents

Think bigger	3
Before you start: Getting everyone on board	5
Step 1: Facilitate collaboration across the organization	8
Step 2: Build your team	13
Step 3: Set up the right data pipelines	15
Step 4: Choose the best use cases	18
Step 5: Get the tools in place to enhance data and build models	20
Step 6: Find the right external data sources	23
Final thoughts: Experiment, test, refine	25
About Explorium	26

Think bigger

To compete today, you need to be able to confidently predict how you'll perform tomorrow. It's simply no longer enough to know how well your organization performed in the past. You need to know what's coming so that you can figure out how to respond to market pressures and volatility, new trends, tastes, and technologies.

These insights can't be extracted from analyzing your historical and current datasets alone. How could they? We're talking about scenarios that are either hypothetical or so new that you don't have a blueprint to help you navigate and understand them. The unprecedented situation we're all living in right now, as the whole globe adapts to a new economic landscape and reality, means you need to think bigger than basic data analysis.

In other words, you need datasets drawn from inside and outside your organization, and you need data science to make sense of it. This is great if you have your own in-house team of data scientists. But, what if you don't have any machine learning experts in your organization? How do you rise to the challenge?

Over the next 26 pages, we'll explain how you can enable data science in your organization with the resources you have available.

We'll show you how to use cutting-edge technology to make the most of your budget and expertise, enabling a productive, machine-learning-backed strategy.

We'll demonstrate how a solid foundation of data science will keep you competitive and successful in the challenging months and years ahead.

Before you start: Getting everyone on board

Even if you've never incorporated sophisticated forms of data science into your operations before, you no doubt already have systems for data collection, management, and analytics in place.

Even in the best of times, no business leader wants to risk wasting resources on big projects that won't deliver results. It's up to you to demonstrate, first, how data analytics have already helped to improve visibility and drive best practices in your organization.

Highlight your data-driven successes to date, especially those with tangible business benefits. Focus on examples of data analysis efforts that informed decision-making in ways that helped boost efficiency, created demonstrable cost savings, increased sales income, opened up new revenue streams, and so on.

However, you also need to emphasize the intrinsic limitations of these efforts. In particular, how could you have gone further with predictive models? How will these data analysis activities be threatened or rendered irrelevant by the tectonic shift in the global economy?

...

Even in the best of times, no business leader wants to risk wasting resources on big projects that won't deliver results. It's up to you to demonstrate, first, how data analytics have already helped to improve visibility and drive best practices in your organization.

When highlighting the benefits of moving towards machine learning, explain how this approach will:

Protect the business by giving your colleagues time and knowledge to prepare and strategize for fast-emerging threats.

Help you identify new and alternative revenue opportunities at a time when core business operations may be hindered.

Give your organization the gift of certainty at a volatile time, driving better decisions and avoiding costly mistakes.

It's worth pointing out, too, that you plan to work with the expertise you already have in-house, increasing machine learning capability through carefully chosen technologies rather than expensive data science salaries (at least, for now).

Once you've laid the groundwork by securing the interest of your C-Suite, you're ready to begin in earnest.

Step 1

Facilitate collaboration across the organization

For any data science project to work, you need to get people on board from all different units of the organization, drawing on their skills, perspectives, and experience. That way, you will combine relevant domain expertise, and the most useful technical skills and ideas, with a clear idea of what different sectors of the business want and need from your data science initiative.

The key here is multidisciplinary collaboration. You need to engage heads of departments to figure out what is required and how this would fit in with their existing processes. You need to engage IT professionals to establish what is possible and how this will work with source systems and target systems, ensuring smooth access to the data you need. You also need to get the blessing of your CEO and CTO.

...

For any data science project to work, you need to get people on board from all different units of the organization, drawing on their skills, perspectives, and experience.

...

Machine learning is a tool
for getting answers, but
it won't deliver results
unless you know how to
use it.

Leaving out any of these voices from your initial scoping and research will inevitably create unnecessary roadblocks and resistance later on. People are scared of change, but only when they feel it's being pushed on them. The more excited you get your colleagues about how this will make their jobs easier and more effective, the smoother your project will be.

Hone your storytelling skills

A huge factor here is effective storytelling. You need to be able to present a clear, transparent narrative to stakeholders that they appreciate and understand. This will underpin their willingness to adopt and accept any future machine learning project — or to dedicate time and resources to your data science planning at all.

It's important to be realistic. Demonstrate the value data science offers to each part of the business, but don't wildly overstate its power. There is so much hype around AI already, and while some are well-founded, it's not a miracle drug. Machine learning is a tool for getting answers, but it won't deliver results unless you know how to use it.

Since you're tackling this yourself, without a dedicated data science team, there will inevitably be some trial and error involved until you perfect the strategy.

Over-promising simply means people will get disappointed and lose interest at the first setback.

Step 2

Build your team

Your machine learning project won't get off the ground unless there are specific people in the organization taking ownership of it. That means you need to put together a dedicated team. This includes people on the business and domain knowledge side to keep things on track, as well as those with the practical roles and skills needed to deal with data and deploying models.

Who should be on your team?

Exactly who you collaborate with depends on the structure, nature, and needs of your business, but here are a few key figures to consider:

IT teams

Your IT department can help you match up datasets with data science tools and platforms (more on that later), get your models production-ready, and ultimately deploy your machine learning models.

From a technical point of view, these are your most important partners.

Leaving them out will create untold integration and operations management headaches, leading to inefficiency and delays.

Data management teams

These are the gatekeepers for acquiring and preparing your in-house data. Involving everyone involved in data management will help you spot important considerations when ingesting data into your machine learning models.

Citizen data scientists

Citizen data scientists are people who don't actually work primarily in data analytics or statistics, but who nevertheless spend time building, or are interested in creating predictive and prescriptive models. In other words, data science hobbyists rather than data science professionals per se. If you have people like this already in your organization, they can prove extremely helpful with more routine data science tasks, provided that you have augmented data discovery and preparation tools, or augmented and automated machine learning tools, in place.

Step 3

Set up the right data pipelines

To make this project work, you need to figure out how to bring data from all over the organization together into a single source of truth, treating it as a coherent resource.

That means moving from a silo mentality to a platform one, which will permit you to collaborate across the organization and scale your machine learning efforts.

Silos are a major barrier to data science. They make it difficult to optimize, transform, and integrate data. If you use many different siloed datasets, you will find that these are riddled with inconsistencies that create a huge harmonization challenge before you even start working with the data they contain.

Switching to a platform approach, on the other hand, means you can combine and clean that data far more easily and efficiently. In fact, depending on how you manage it, you may be able to automate much of the heavy lifting, getting your project off the ground much faster.

It's important to remember that, in addition to moving to a literal platform in the sense that you need centralized technology, you also need to get the whole organization thinking in this way. Every team needs to view their data as a collective resource of value to other departments and sectors across the business, and not a proprietary resource they hide away for their own, exclusive use.

...

Every team needs to view their data as a collective resource of value to other departments and sectors across the business, and not a proprietary resource they hide away for their own, exclusive use.

Step 4

Choose the best use cases

Start with a strong but relatively simple use case to test your approach. It's best to pick something you plan to use internally during this experimentation phase, rather than a customer-facing project or product.

This will take the pressure off you to get it absolutely perfect on the first attempt. It also means your colleagues can get used to trying out the technology and seeing how it benefits them before they are expected to sell the idea to people outside the company.

Don't be too ambitious, at least not yet. Consider the budget and resources that are available to you and the timeframe you have for the project. You don't want this to drag on for too long.

When brainstorming use cases, it's wise to focus on those that demonstrate a way of monetizing your data. That might mean creating a machine learning-driven product or service that will ultimately turn a profit, but it may also mean:

- Streamlining a particular process or operation, which in turn reduces costs or makes you more efficient.
- Boosting internal and/or external communication and information-exchange, for example, to make the business more productive or to improve customer service, loyalty, and retention.
- Using corporate data assets to predict or suggest opportunities to gain market advantage.

For instance, one simple use case to get you started might be to create a chatbot that uses natural language processing to answer internal customer support questions, or which you roll out to new recruits during training to cover frequent queries.

Once you have proof of concept, you could then replicate this for use in customer service, reducing the pressure on your team, or allowing you to remain available to customer queries even outside of working hours.

Step 5

Get the tools in place to enhance data and build models

Since you don't have data scientists in the organization, you will need to ensure you have the right tools and platforms to help you build the models and other machine learning applications you want to bring to your business.

There are all kinds of machine learning tools and software you can try out that won't demand coding and programming knowledge. Some of these offer a complete data-science-as-a-service solution, which does all the heavy lifting for you.

Look out for platforms that reduce your workload, rather than adding to it. For example, Explorium combines machine learning with data enrichment. This means you can use it to generate extra features that give you more options for analysis and more ways to model data, all

...

There are all kinds of machine learning tools and software you can try out that won't demand coding and programming knowledge. Some of these offer a complete data-science-as-a-service solution, which does all the heavy lifting for you.

from the same datasets. It also automates the process of cleaning up and harmonizing data so that it's ready to use, allows you to connect to thousands of external data sources for contextualization, identifies the best algorithms for your data, and provides interpretation tools to help you uncover patterns and insights quickly and easily.

Step 6

Find the right external data sources

To really get a competitive edge, it's likely that you will also need to look outside your own organization for valuable, up-to-the-minute data. As we've seen, your historical and current data can only tell you so much about what your customers are saying, what the market is doing and how you can seize on opportunities as they emerge.

However, working with external data can be a challenge. First, you need to worry about regulations such as GDPR, which affect how you can utilize and store user data. New regulations like these are being rolled out all over the world and mean you really do need to trust that the source of the data is up to speed on their privacy and security responsibilities.

Second, working with multiple data sources is a much larger version of the problems you have in-house when working with siloed data. You may be dealing with a hundred different formatting styles. You may

find that the quality and accuracy of datasets vary widely from source to source. You may find, in short, that you spend more time preparing that data for use in your models than you do creating and refining the models itself.

The best way to get around that is, again, to embrace a platform approach. By using a robust platform designed to connect to a wide range of pre-vetted external data sources, you outsource many of these headaches.

What's more, the best of these will do a lot of the data cleaning and processing work for you, so that you can hit the ground running rather than spending weeks trawling through data and tidying it up before you can use it.

Final thoughts: Experiment, test, refine

If you're new to data science, you will find that much of learning happens through doing. You need to experiment with different models, different datasets, and different ways of combining data from inside and outside the organization.

That said, your experiments need to be aimed towards a goal. Make sure you're continually learning from them, documenting what you've found, what works, what doesn't, and how you can improve for the next attempt. These will help you refine your projects and ultimately to scale them up.

Experimenting and adapting is far easier when you have the right tools to facilitate this. Getting a robust platform that removes a lot of the data preparation work, as well as automating the process of applying algorithms and generating models, means you have far more time to concentrate on what matters. Most importantly, you can automate wherever you can and spend your energies on honing the data science strategy and products that will allow your organization to flourish.



About Explorium

Our automated data discovery and feature generation platform automatically connects a company's data to thousands of relevant premium, partner, and open data sources to extract an optimal feature set based on model impact. We're creating a new category as the first company to empower and service business leaders and data scientists with end-to-end automation of data discovery and feature generation —

fueling superior decision-making and driving real business impact.