Enterprise AI for Business Preparedness

dataiku

How to Become More Agile in an Era of Economic Change

WHITE PAPER

www.dataiku.com

Introduction

Disruption defines our world, and the latest health crisis is just the most recent example. Whether it's upheaval brought on by new technology, economic downturn, environmental disaster, or new competitive business development, the unfortunate reality is that most businesses are not set up to be able to quickly and seamlessly pivot their operations to effectively cope with big change.

Though spending on AI initiatives has increased exponentially in recent years, it has remained tangential to many companies' central operations, viewed more as an experiment than an indispensable organizational asset. However, in the second half of 2020 and beyond, this paradigm is poised to shift as Enterprise AI becomes a critical component of companies' strategy to recover from crisis and bring more preparedness for the future via AI systems that are persistent and resilient.

"Schneider is more than 180 years old. We started in iron and steel and now we are digital solutions for energy, using automation. What we've learned is that there are two kinds of pivots: the pivots you initiate, which are good for the company, and the pivots the environment imposes on you. Those are very painful. So our obsession in the past 20 years has been to anticipate, to choose our pivots and transform continuously as a company."

Jean-Pascal Tricoire, Chairman and CEO, Schneider Electric[1]

This white paper will go through several AI use cases that businesses can start right now in order to help in a time of economic change. It will also cover mid- to longer-term strategies for AI around reuse and model maintenance to help ensure preparedness for future disruption and challenges.

Throughout, there are a few key points to keep in mind:

- AI isn't a magic bullet, and it won't make businesses immune to the world's changes.
- However, robust AI implementation provides the accelerator necessary to both cope with and emerge stronger from change.
- The goal of Enterprise AI has been — and should continue to be — to enhance (not replace) people, allowing them to make better decisions and be more efficient, etc. AI can only be a change management tool if this remains the case.

[1] https://www.accenture.com/fi-en/insights/consulting/business-disruption-innovation

AI Use Cases to Start Now

In an ideal world, organizations would already have fundamental AI use cases in place that allow the business to optimize costs and accelerate their ability to execute on critical business functions, and they would not wait until they are already in the midst of a crisis to put these in place. For many reading this white paper, this may not be the reality; therefore, this section will unpack some of these critical use cases on which companies can start now.

Optimization

For those looking to kickstart their path to Enterprise AI in a turbulent period, finding areas of optimization (i.e., where the business can do more with less) is a good place to start, as these use cases offer clear business objectives and tangible return on investment (ROI). Every organization, no matter what industry, has areas that can be optimized by AI processes. For example:

- Optimizing marketing spend with machine learning-powered attribution for hyper-personalization.
- Using machine learning to consider factors like weather, traffic, and a myriad of other constantly-shifting data sources to optimize workforce allocation.
- Moving toward machine learning-powered predictive maintenance to decrease downtime, increase quality, and improve scheduling.

For a more concrete and real-world example, a Dataiku customer who is a major provider of travel and airfare vacation packages had mostly revenue-related use cases in their pipeline in late 2019 and early 2020. Now, they have pivoted to focus on cost-effect, bottom-line cost optimization initiatives (for example, an airplane meal forecasting model for multimillion-dollar savings).

Prioritizing Time-to-Value

Whether the business decides to tackle optimization or acceleration use cases, one critical component to keep in mind is time-to-value. More than ever before, given the current economic climate, organizations must prioritize use cases that can start delivering results in a matter of weeks, not months or years.

Successfully prioritizing time-to-value involves a combination of choosing the right use cases (i.e., large enough for big impact, but not so expansive that they'll be slow to get buy-in and implement) and also the right tools that help reduce friction and repetition on the path to Enterprise AI. Data science, machine learning, and AI platforms (like Dataiku) provide this experience, reducing the time it takes to connect to and clean data, create machine learning models, and deploy them to production.

Dataiku was named a Leader in the Gartner 2020 Magic Quadrant for Data Science and Machine-Learning Platforms and scored highest in two of the four use cases in the Gartner 2020 Critical Capabilities for Data Science and Machine-Learning Platforms.

Of course, these are just a few examples, and each business will need to find its own areas of optimization. Some key questions to ask while choosing optimization use cases are:

- How will this use case specifically allow for cost optimization, and how will this be measured? Be precise.
- Where will the data come from, and does it already exist? Avoid use cases where you're starting from square zero and will have a high time-to-value.
- Is the use case a heavy-hitting application that really provides value at scale? Choose use cases large enough to have an impact, but not so large that they will take many months to build.
- Will using machine learning or AI processes prove larger value or ROI than existing processes? If the answer is "no," choose another use case.

Acceleration

AI use cases that accelerate the business's ability to execute on critical business functions are also optimal for getting started (especially during periods of economic change). These use cases include things like:

- Machine learning-based demand forecasting, as older modeling techniques fail to incorporate the wide variety of data sources needed to produce results precise enough for the modern enterprise and a changing environment.
- Churn prediction and prevention using more advanced modeling techniques not only to identify potential churners, but predict who is most likely to respond positively to marketing efforts to stay.
- Prioritizing sales efforts by leveraging machine learning to determine who is most likely to buy today, allowing teams to accelerate by targeting their energy.
- Having a robust, audit-ready environment to support critical models, accelerating the ability of highly regulated industries (financial services, pharmaceuticals, energy, etc.) to respond even in tumultuous periods.

Operationalizing for ROI

Companies are realizing now (perhaps with an urgency that didn't exist before) the importance of operationalization and getting models out of the lab and into production. For example, one Dataiku customer — a semiconductor firm — that was pursuing production line optimization use cases with factory sensor data in late 2019 and early 2020 has now significantly accelerated its efforts to get these use cases into production.

The ability to operationalize and quickly deploy machine learning models to production is always important, but even more so in times of change. Why? For those just getting started, it's the only way to realize ROI and actually start optimizing costs, increasing revenue, and seeing change. For businesses that already have models in production, the underlying data has fundamentally shifted, and slow operationalization capabilities mean a longer time to deploy new models that better fit today's (or tomorrow's) reality.

Accelerating by Enabling Remote Work

One of the challenges to ramping up AI efforts quickly across the enterprise is remote distribution. Whether teams are distributed during normal times or exceptional circumstances require everyone working from home, remote work is the ultimate litmus test of the data organization's robustness.

Many inefficiencies may go unnoticed when working in the office that ultimately lead to significant loss of time or of project relevance, but these issues can at least be partially mitigated by informal discussions and water cooler interactions.

The Challenges of Remote Data Science

A data team set up to be efficient remotely opens new opportunities for productivity; however, it also brings challenges in and of itself, namely:

- **Access to systems:** Connection to underlying data systems may prove challenging in a work-from-home environment. Whether accessing the various data sources or the computational capabilities, doing it in a remote setting is often challenging.
- **Collaboration within teams:** Data projects are rarely one-person jobs, and the easiest way to get teams working together is to get them to sit together. Without this physical proximity, individuals are often siloed in their execution of projects.
- **Collaboration across teams:** Data projects are not only about data, but also require strong involvement from business teams to build experience, generate buy-in, and validate relevance. They also require data engineering and other teams to help with the operationalization steps. While watercooler discussions play a pivotal part in the (often limited) success of cross-team collaboration, a full work-from-home setup may prove highly disruptive to execution.
- **Reuse over time:** Capitalizing on past projects is essential to maintain productivity. However, capitalization within large code repositories often manifests itself informally, and the lack of off-the-cuff discussions may significantly limit the ability to reuse past work.

How to Actually Make It Work… Remotely

The right data science, machine learning, and AI platforms enable remote work at their core by addressing all of these challenges. For example, Dataiku allows people across the organization to access all data and work together on projects in a central location, facilitating good data governance practices combined with widespread vertical (e.g., data scientist to data scientist) as well as horizontal (e.g., analyst to data scientist or business user) collaboration.

During the current health crisis, large, multinational Dataiku clients (including Unilever, GE, Pfizer, and more) have avoided machine learning operations disruptions due to their strong remote culture and global distribution — they simply use VPNs and web browsers to access data, projects, and colleagues via Dataiku.

Continued from page 4

The reality is that data science, machine learning, and AI are exactly what allow companies to be more efficient across all areas of the business, so these initiatives are perhaps more relevant now than before. This applies to businesses just getting started as well as those who have already invested heavily in AI initiatives.

To take one final example, a global manufacturer and Dataiku customer unfortunately had to lay off 10% of its workforce following the 2020 crisis, including some data engineers and other staff supporting data science and machine learning efforts. However, they are still supporting continuity on both “critical” (i.e., impacting $20-$30 million) and “critical mass” AI projects (i.e., those deeply embedded and touching multiple aspects of the business).

Despite the cuts, this company continues to invest heavily in advanced analytics projects and retain key analytics personnel because they see data science, machine learning, and AI efforts as an organizational asset, business critical to emerging from the crisis stronger.

The Cost of AI and the Critical Role of Reuse

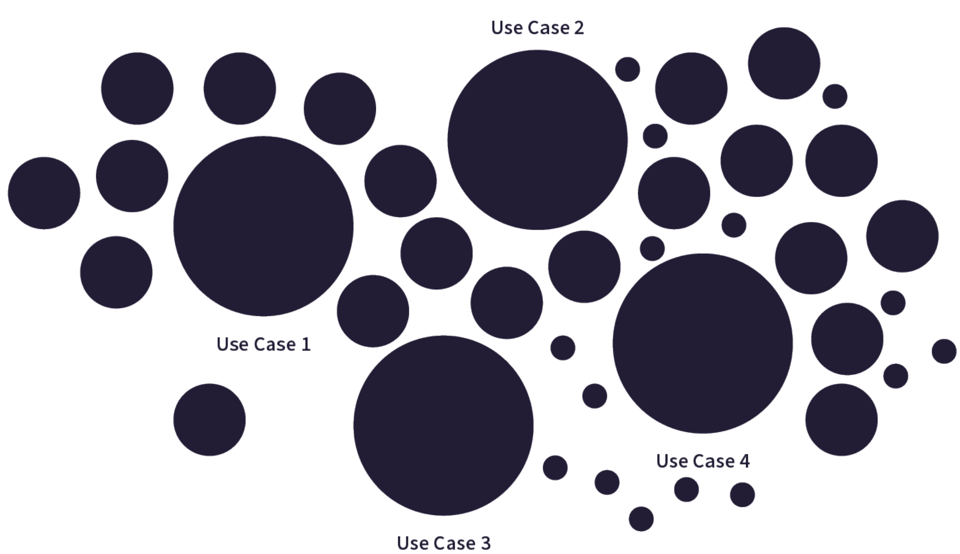
It would be naïve to ignore the fact that data science, machine learning, and AI initiatives represent a cost of their own. How, in a time where cost optimization is more critical than ever, can businesses — whether just getting started or looking to improve efficiency — manage and justify the cost of AI itself?

Of course, there are obvious, tangible costs (like that of tools and technology), which should certainly be audited in order to successfully scale. Strategies include minimizing the overall number of tools — investing in one truly end-to-end platform instead of piecing together multiple tools for ETL, model building, operationalization, etc. — and to embrace elastic resources when it comes to data computation.

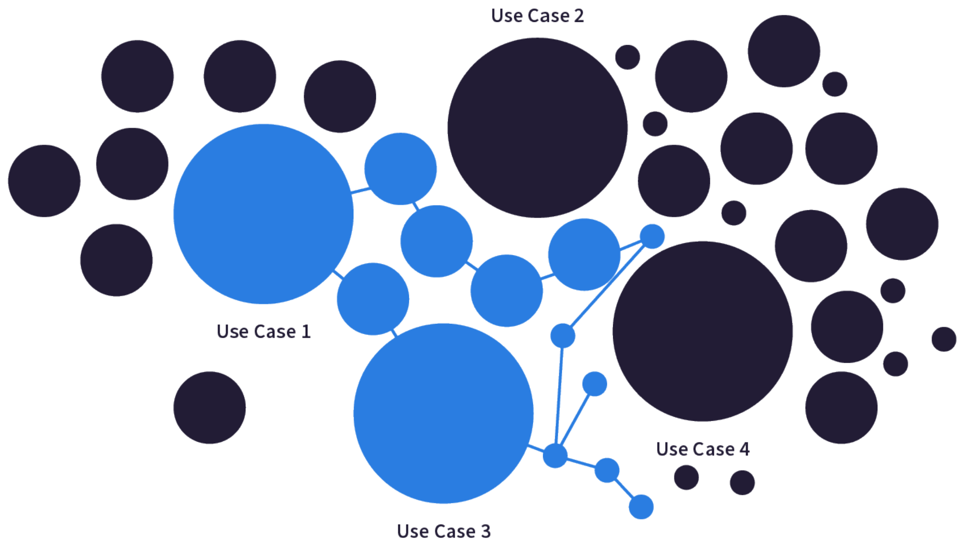
But there are also many less tangible costs that tend to bog down organizations' efforts by adding up over time, hampering their ability to scale and profit from Enterprise AI, including data cleaning, preparation, operationalizing, and more. Common sense and economics tell us not to start from scratch every time, and that is exactly the principal behind reducing costs associated with Enterprise AI.

Reuse is the simple concept of avoiding rework in AI projects, from small details (like code snippets that can be shared to speed up data preparation) to the macro level (like ensuring two data scientists from different parts of the company aren't working on the same project). Capitalization in Enterprise AI takes reuse to another level — it's about sharing the cost incurred from an initial AI project (most commonly the cost of finding, cleaning, and preparing data) across other projects, resulting in many use cases for the price of one, so to speak.

But how exactly do reuse and capitalization ensure scale? By increasing the number of use cases addressed with AI projects while reducing the impact of the costs outlined above. For example, say the business has a list of four uses cases in mind to start their Enterprise AI efforts. However, in addition to these four, of course, there are lots of other potential use cases across the business:



Capitalization means that while tackling these larger, high-priority use cases, the organization can also take on lots of other smaller use cases by reusing bits and pieces, eliminating the need to reinvent the wheel with data cleaning and prep, operationalization, etc.



Capitalization and reuse sound easy in principal, but in practice, they require strong, enterprise-wide, centralized processes where:

- People can easily access information, including who is working on what projects.
- People can transparently consume things done by others (including seeing data transformation, models, etc.)
- People can take, reuse, and adapt work done by others.
- Data experts can capitalize (and monitor) a portfolio of data treatments to be used across the organization.
- Data experts can easily build and share projects to be used by others.
- People — whether coders or not — can work efficiently in their preferred way.
- Data leaders can ensure the quality of AI projects, ensuring that capitalization and reuse are being used properly

Model Maintenance Strategies

Building an Enterprise AI strategy that is fit to carry the business through economic highs and lows isn't just about implementing a list of use cases and leveraging reuse to expand on those use cases. It's also pivotal to have systems for monitoring models in production and to be able to quickly introduce, test, train, and implement new models in order to shift strategies or adapt to changing environments on a dime.

model drift
/ˈmɒd(ə)l/ /drɪft/

noun, verb

When a model deployed in a production environment deteriorates and is no longer fit for purpose. This can occur either because of:

1. Data drift, meaning the data on which the model is based has shifted over time.
2. Concept drift, meaning the statistical properties of the target variable (which the model is trying to predict) change over time in unforeseen ways.

Related: model decay, model failure

Short-Term: Addressing Model Drift During a Crisis

For a sense of just how important model monitoring can be to a business, consider the fact that regional lockdowns in March 2020 brought some chaotic days across all industries, including retailers where every day started to look like a Sunday, disrupting normal processes and operations. Since then there has been some stabilization, but it's likely that disruption will happen again once lockdowns get lifted and human behavior changes again, which likely means models drifting not once, but twice.

Organizations often need to reframe data science projects quickly during turbulent times, as many of the assumptions underlying existing machine learning projects are no longer valid. Some key questions to answer for businesses that have models in production during periods of rapid change are:

- Have the business needs related to the models changed given the circumstances?
- Has the availability, reliability, and relevance of input data changed? If so, can new data sources be more relevant and better connected to the prediction targets?
- Is the granularity of the predictions or of the models (e.g., the time horizon for forecasts) still adequate?
- Should the predictions be consumed in a different manner during a volatile period? For example, if the predictions used to be directly used by a larger automated system, should a human expert be added in the loop?

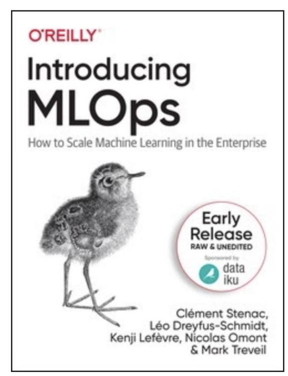
In general, models trained on historical data should be assessed critically during periods of dramatic change. This is valid both for models trained before the period of change itself but also for models developed during and after volatile economic conditions, as the reality might be different than before.

The way to address model drift depends on whether it is possible to get immediate or delayed feedback. In both cases, the efforts related to model monitoring should be increased during and after times of change given the high risk of model drift (for example, increasing the frequency of checks and the resources devoted to investigating alerts):

Situation	Example	Drift Detection Mechanism
Immediate feedback	A recommender system in an e-commerce setting	Direct comparison of the prediction and the new measurement
Delayed feedback	A forecast for next month's sales	Detection of a potential change of the input data distribution or of the predictions distribution Comparison of the prediction and the new measurement when the latter is available

Long-Term: Developing Strong MLOps Practices

At its core, MLOps is the standardization and streamlining of machine learning lifecycle management. MLOps isn't just important because it helps mitigate the risk of machine learning models in production, but it is also an essential component to scaling machine learning efforts (and in turn benefiting from the corresponding economies of scale). Going from one or a handful of models in production to tens, hundreds, or thousands that have a positive business impact and can be easily shifted in the event of a crisis or shift will require MLOps discipline.



Good MLOps practices will help teams at a minimum:

- Keep track of versioning, especially with experiments in the design phase.
- Understand if retrained models are better than the previous versions (and promoting models to production that are performing better).
- Ensure (at defined periods — daily, monthly, etc.) that model performance across the organization is not degrading in production.

Overall, model monitoring (both short- and longer-term) highlights the need for better overall transparency into the machine learning project lifecycle for all, including managers and executives, to make faster business decisions. The more transparency that can be built into AI processes at all levels, the better prepared the organization can be to leverage those AI systems to make critical, data-driven decisions in unstable times.

Conclusion and Next Steps

Unfortunately, some businesses are coming out of the 2020 health crisis with negativity around AI and what it can bring to decision making at companies. No one was able to predict what happened, meaning it's obviously impossible to predict the future; so what's the point in investing in predictive analytics initiatives, including data science, machine learning, and AI?

Ultimately, the goal of many AI systems is not to predict the future with 100% certainty, but to model and prepare for different scenarios. For example, Dataiku has lots of customers who are pursuing prescriptive analytics — like what-if scenarios and Monte Carlo simulations — around different hypotheses (e.g., insurance firms running simulations on the insured population's health habits during lockdown to understand the post-lockdown insurance landscape).

For companies that survive, AI will become more important than ever as an organizational asset for handling large-scale change with greater ease. No matter where organizations are on their AI journey, post-2020 should be a time to;

1. Reassess AI use cases across the board, looking for those that bring the most ROI and value to the company. See The ROI Toolkit for Data Science, Machine Learning, and AI Initiatives.
2. Reduce costs surrounding AI initiatives by embarking on a campaign for reuse and capitalization in data projects. See The Economics of AI.
3. Ensure robust model maintenance strategies, encouraging transparency on model performance throughout the enterprise. See Introducing MLOps.

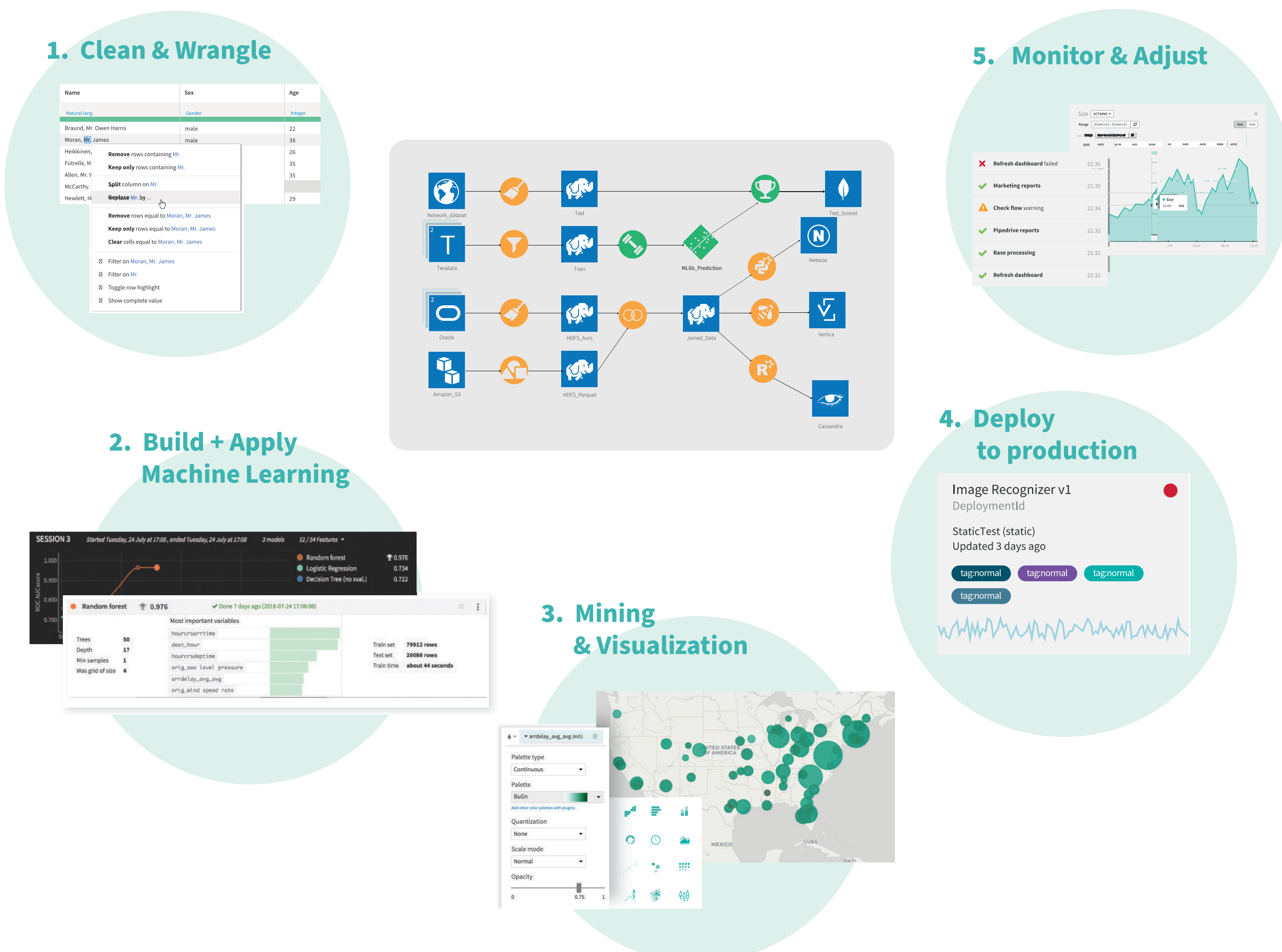
Your Path to **Enterprise AI**

Dataiku is one of the world's leading AI and machine learning platforms, supporting agility in organizations' data efforts via collaborative, elastic, and responsible AI, all at enterprise scale. Hundreds of companies use Dataiku to underpin their essential business operations and ensure they stay relevant in a changing world.

300+
CUSTOMERS

30,000+
ACTIVE USERS

*data scientists, analysts, engineers, & more



WHITE PAPER

www.dataiku.com