GUIDEBOOK

Accelerating AI Maturity

A Guide to Reducing Costs and Creating Value With AI Applications

Introduction

According to a NewVantage Partners executive survey, **over 90% of business leaders report** that challenges to becoming data-driven are people and business process related and not about technology[1]. Resultantly, while technology still plays a role, there are many other facets for organizations aiming to level up their AI maturity to consider. These facets, whether conspicuous or not, ultimately end up playing part in the cost optimization and value creation associated with AI applications.

> ***"Maturity is the focus on planning for and adapting to either an incremental or transformational process change. It is measured not only by cost reduction, but also by reduced cycle times, lower error rates, scalability, and business agility (the ability to respond or demonstrate insights previously unavailable)."***
>
> *- Gartner, Artificial Intelligence Maturity Model*[2]

To seamlessly transition from one wave of AI to the next, organizations need an acute understanding of their AI maturity. The exercise can enable them to benchmark any potential growth toward mastery of AI (and identify if their AI is acting as a utility, business enabler, or business driver) alone and versus competitors, to strategically plan which internal organizational steps need to be taken to reach their goals for AI, and/or communicate that vision back to key stakeholders to measure success. Therefore, assessing and tracking AI maturity is pivotal for any organization interested in accelerating their journey toward Enterprise AI and extracting more value out of their investments.

[1] http://newvantage.com/wp-content/uploads/2020/01/NewVantage-Partners-Big-Data-and-AI-Executive-Survey-2020-1.pdf
[2] Gartner "Artificial Intelligence Maturity Model," Svetlana Sicular, et al, 18 March 2020. https://www.gartner.com/document/3982174 (Gartner subscription required)

Dataiku's AI maturity model details a five-step journey for organizations interested in improving their AI maturity:

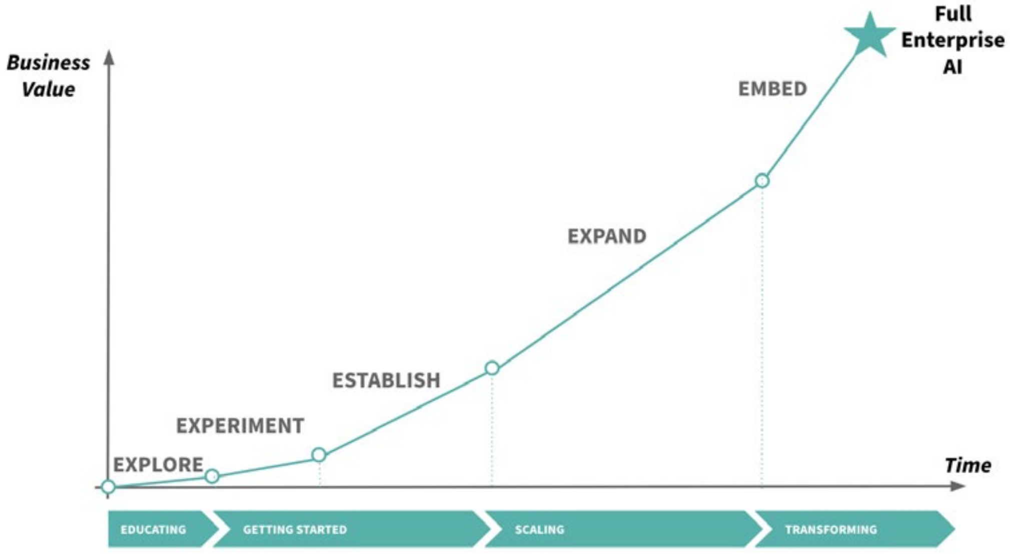
- **Explore:** Explore what AI is and means for the organization, evangelize the need to leverage AI, and find early adopters.
- **Experiment:** Experiment with the value of AI with first projects and build awareness.
- **Establish:** Establish tangible value from a few initial use cases and lay the foundations to scale.
- **Expand:** Expand usage of AI across the organization and accelerate business value, building on foundations previously laid out to spread to all departments and functions of the organization.
- **Embed:** Embed AI in every single activity so that AI is part of the DNA of the organization and wholly merged with overall strategy.

After everything is embedded, the organization enters into **full Enterprise AI**. This is the ability to embed AI methodology — which combines human capacities for learning, perception, and interaction all at a level of complexity that ultimately supersedes our own abilities — into the very core of an organization's data strategy.

It is important to note that achieving full Enterprise AI doesn't just mean achieving the methodology. Rather, AI is everywhere, inseparable from an organization's global strategy for achieving its most critical business objectives. AI will be created (or at the very least, leveraged one way or another) by everyone within the organization and, ultimately, will impact every process.

Dataiku AI Maturity Model

A 5-step journey towards your organization's AI maturity



It is important to note that our AI maturity model is meant to equip data and analytics leaders with a guiding framework for their AI strategy. It can be used to demonstrate the value of analytics and build momentum internally, a key piece of the puzzle when trying to transform an organization at scale. Not all organizations should strive to reach the top level of maturity at first, but rather start at the level appropriate to their bespoke business objectives and think about moving forward in stages — a process that will take more time for some organizations versus others.

One thing that Dataiku's AI maturity model has in common with that of industry research entities is that, in recent years, AI adoption has targeted "low-hanging fruit" projects (i.e., those we see in our Establish stage) with high value and an obvious Return on Investment (ROI). An example of this might be a pharmaceutical firm that has hundreds of active clinical trials that cost millions of dollars. The "low-hanging fruit" project may not necessarily be scientifically complex, but can end up saving the business hundreds of millions of dollars.

Moving forward, though, most of the benefits of AI will be provided by applications that are less obvious and potentially more costly to implement and deploy within organizations. The main challenges organizations will face as they move along their AI maturity journey is reducing the cost of building and operating AI projects.

This white paper will:

- Provide a light framework for getting past these "low-hanging fruit" projects and on to the next wave of AI (in Dataiku terms it's to the "Expand" phase and beyond)
- Detail key challenges on the road to truly pervasive AI (and ways data leaders can guide their teams in the right direction)
- Explain how a collaborative data science tool can help reduce costs associated with data projects
- Highlight concepts and strategies like capitalization, reuse, and MLOps and their role in ushering organizations to the next wave of AI

Challenges Remain on the Road to Pervasive AI

In order to reach the Expand and Embed stages of AI maturity, organizations need to critically analyze the steps to **pervasive AI**, where more employees in an enterprise will benefit from AI via augmented applications, moving AI and machine learning capabilities closer to those taking action.

Making sure people work together to maximize output and shared knowledge is paramount — getting people to connect can and should be done directly within the tools data science teams are using and embedded in their workflows. In each section below, data science tools already reduce part of the cost of AI projects, but challenges persist.

Data

While data science tools can help reduce costs via tasks like data sourcing, simplifying and improving aspects of data quality, data labeling, and connectivity, one may still ask how to get the right data for the analytics needs, as it is presented in business applications. Many data leaders (think CDOs) are beginning to embed "business translators" into the organization to effectively translate business needs into data needs to make AI pervasive.

Regarding the business translator, Gartner states "this could be a business-savvy data scientist or citizen data scientist, an analytically minded business person or a process engineer (process modelers or business analysts focused on process design) who is mindful of business optimization opportunities derived from analytical assets."[3] By injecting these translators (who evolve into trusted figureheads across data teams) into the appropriate pockets of the business to identify data requirements, oversee data workstreams, and act as mediators and go-to points of contact for members of the development and operationalization teams as well as to executive stakeholders, organizations will be able to infuse more agility into the transformation and delivery required of large-scale data projects.

[3] Gartner "Use 3 MLOps Organizational Practices to Successfully Deliver Machine Learning Results," Shubhangi Vashisth, et al, 2 July 2020. https://www.gartner.com/document/3987106 (Gartner subscription required)

Operationalization

By providing pre-built frameworks for moving models into production, data science tools can reduce costs associated with model maintenance, monitoring, and robustification. While there's no denying that industrializing and extracting actual value from AI projects is hard (due to both technical and people-related challenges), there are many projects that are running live in organizations today because they have been "robustified."

Dataiku surveyed over 200 IT executives asking "On average, how long does it take to release the first version of a machine learning model in production?" Over half cited between three and six months, representing a massive cost in labor and lost revenue for the amount of time the model is not in production and able to drive value for the business. There's also the opportunity cost to consider — when all the focus is on pushing one model into production, it takes teams away from dedicating time to new projects.

Robustifying a model consists of taking a prototype and preparing it so that it can actually serve the amount of users in question, which often requires a significant amount of work, typically from data engineers. In many cases, the entire model needs to be re-coded in a language suitable for the architecture in place, often causing delays in deployment. Once this is complete, it has to be integrated into the company's IT architecture. Even further, the right people need to access data where it sits in production, a process often made more difficult due to technical and/or organizational data silos.

How then, can teams build robust feedback loops with input from everyday users of the augmented applications? The answer is via a sound MLOps strategy. MLOps takes operationalization a step further, encompassing not just the push to a production environment but the maintenance of those models (and the entire data pipeline) in production.

Once a model has been operationalized and its performance begins to degrade, an update can be triggered by the data scientist. This typically involves either retraining the model with new labeled data or developing a new model with additional features. Either way, the goal is to ensure that the business is not negatively impacted. At a time when issues like responsibility and bias are at the forefront, MLOps becomes even more vital to close the feedback loop between operationalized models and their impact.

We'll touch on the role of MLOps in enhancing AI maturity later in this guide, but for a complete overview of MLOps — including its challenges, personas involved, key features, and more — read ***Introducing MLOps: How to Scale Machine Learning in the Enterprise***[5].

When put into production, a model can start scoring data it has never seen. One can imagine a situation where because the train and validation of the design and train stages has not been set properly, the model in production scores badly.

In Dataiku, users can easily create a validation feedback loop to verify that the newly scored data achieves the original goal of the model. To do this, one creates a new validation set composed of newly labeled data, containing both the score given in production by the model and the final observed values. Then, an evaluation recipe is used to compute the true performance of the "saved model" against this new validation dataset.

Governance

According to the Harvard Business Review, **53% of organizations** have yet to develop a strong, business-wide data governance approac[4]. While today's data ubiquity has empowered organizations to create, store, and leverage data in new ways, this has spiraled into a complex digital environment that desperately requires some level of policy and oversight for its management (and encompasses everything necessary for managing data security, privacy, risk, and regulatory compliance).

As organizations attempt to identify ways to govern and manage their data, many might be asking, "How can I keep a human-in-the-loop approach to meet regulatory constraints and ensure decisions are transparent and auditable?" Whether burdened by the task of managing an ever-increasing volume of data, new security challenges that may arise as more data is shared across the enterprise, or understanding the proper regulatory requirements on a global scale, there are certainly roadblocks associated with data governance. Further, as the number of projects in production increases, visibility into who is using what data, how, in which models, and how these models are deployed is eroded.

[4] https://hbr.org/resources/pdfs/comm/microsoft/MicrosoftDataGov1.27.20.pdf

In order to effectively attain that balance between a human-centric strategy for compliance and ensuring transparency, organizations need to transition from strictly data governance to AI governance. What does this mean? Historically, data governance oversees a range of activities, such as data security, reference and master data management, data quality, data architecture, and metadata management.

Now, with a growing adoption of data science, machine learning, and AI, there are new components that should fall under the data governance umbrella, namely machine learning model management and Responsible AI governance, making data governance a truly multipronged strategy. It's not just about protecting data and defining who is responsible for what, but rather ensuring the right people, processes, and systems are in place to do that.

Times of economic change and uncertainty can ignite massive changes in underlying data, causing machine learning models to degrade or drift more rapidly. Model monitoring, refreshes, and testing are needed to ensure the model performance meets the needs of the business on a continuous basis. Further, when it comes to ensuring decisions are transparent, it comes down to making sure models do what they're intended to do — they're making real-world decisions so having intimate knowledge of the decisions they make and ensuring the models are explainable and traceable is critical for all parties involved. To go deeper on the pillars and pitfalls of data governance, check out this **white paper**.

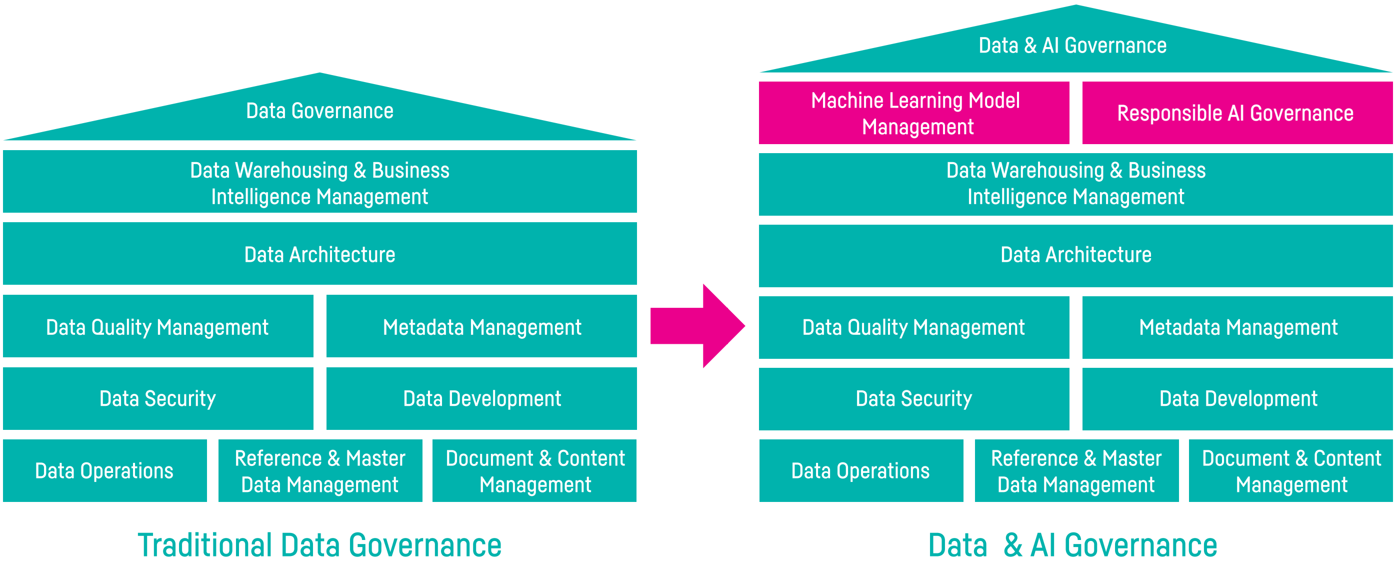


Figure: Moving from Traditional Data Governance to Data & AI Governance

Change Management

To move to the next wave of AI (here, we're referring to the "Expand" phase and beyond), AI must appear everywhere. To scale, the organization has leveraged foundations previously built to empower every department and every subsidiary to deliver AI across the organization. It is important to note that additional drivers (such as strategy, budget, and talent) all play a role in supporting the global spread of AI use and may be more important to consider in order to effectively move from one phase to the next. These drivers of AI maturity are interconnected and need to advance at the same pace or organizations run the risk of being overbalanced. For example, in order to advance staff, the organization will need to delve into more advanced use cases.

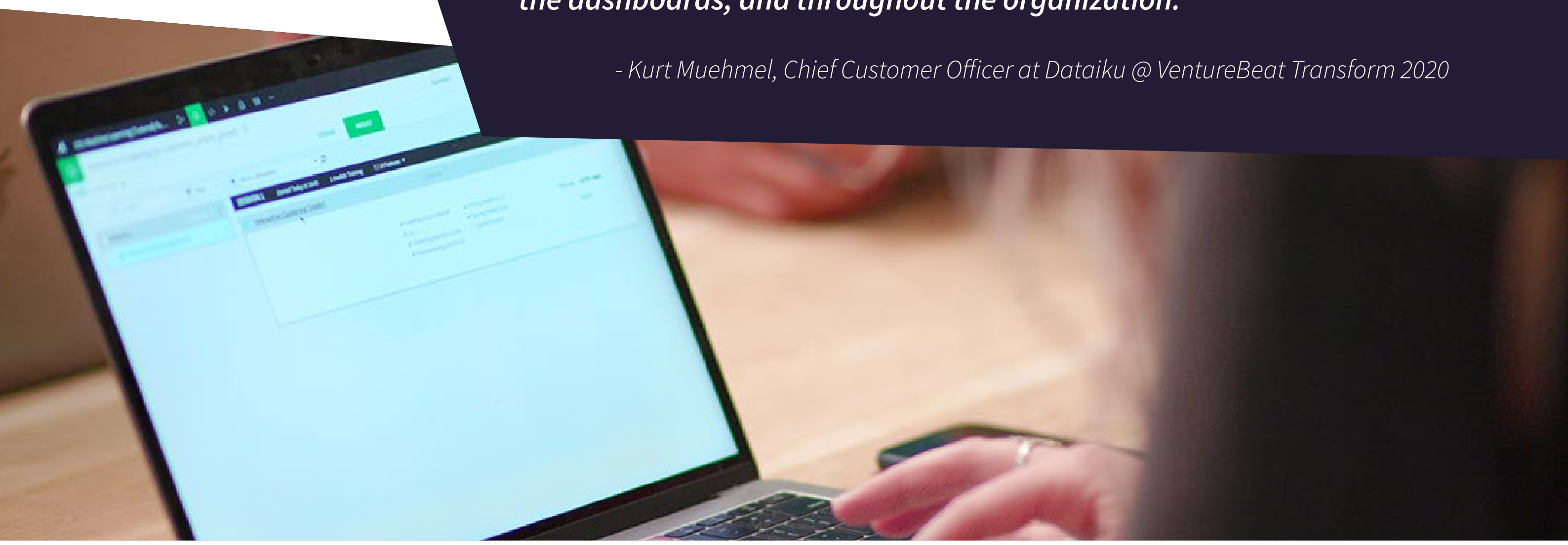
In addition to identifying the right operating model that fits the organization's context, teams need to consider the role of AI in the company's top-down strategy, budget allocation, talent acquisition and retention, enablement, and so on. This is not an overnight process, and teams should be dedicated to lead the way and implement change management. In many instances, organizations may not be open to replacing or changing their existing applications and therefore need to augment their business processes with AI.

To reduce the change and normalize end user training and adoption, there needs to be a culture shift in the way training is viewed. AI experts see training as a need to stay knowledgeable in an ever-evolving technology space, so implementing a strategy on training (not only for onboarding but ongoing) is essential to talent retention. For non-experts, AI is such a shift that it requires putting a lot of time, energy, and resources into it via a personalized, multi-step training, ingrained in the company strategy and culture and inclusive of hard and soft skills. It is the hope that after this training, non-experts can become autonomous on data projects up to the use of standard AI techniques.

To effectively transition from the "Establish" to "Expand" stage of AI maturity (and beyond), teams need a fully functional platform (inclusive of systems, processes, and people training) to be rolled out to all business units. While the usage of AI is expanded and business value accelerated in this new stage, data projects still remain fairly specific, lacking cultural pervasivity. It is not until the "Embed" stage that AI is fully woven into the cultural DNA of the organization.

"After scale, the goal is to get everything embedded, to embed the analytics, to embed the AI processes directly into the applications, into the dashboards, and throughout the organization."

- Kurt Muehmel, Chief Customer Officer at Dataiku @ VentureBeat Transform 2020



Cost Optimization for AI Applications

Particularly on the backdrop of the global health crisis, organizations are beginning to monitor their unique recovery process and analyze the shifting dynamics of the economic landscape in their sector. Data leaders considering implementing a data science platform to help streamline and accelerate these data efforts (and infuse added levels of agility) no longer have the luxury of gradually deducing if the total cost of ownership (TCO) and overall value align with their greater business objectives and financial position.

In order to achieve success using data, **it has to** — and has to be done in a way that is well defined and clearly reduces TCO. We'll give a few examples on how this can be done in the next section.

Capitalization and Reuse for Enterprise AI

Regularly maintaining and monitoring AI projects is a task that cannot be ignored (or at least not without a financial impact). Because data is constantly changing, models can drift over time, causing them to either become less effective or, worse, have negative implications for the business. Further, the more use cases the company takes on, the harder it is for maintenance to be properly addressed, not to mention the rising costs involved.

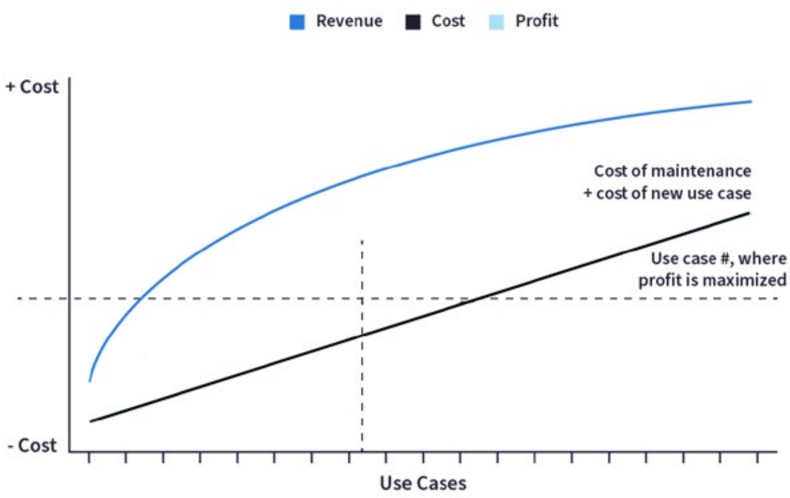
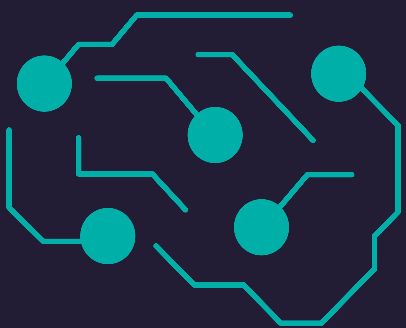


Figure 1: Cumulative revenues, costs, and profits over time (# of use cases); note that after use case #, profit is decreasing due to increased costs and stagnation of revenue.

In order to reduce costs associated with various steps of the data-to-insights process (from data prep to model maintenance), teams can leverage reuse and capitalization. Reuse is the concept of avoiding rework in AI projects, such as making sure two data scientists from different business units aren't duplicating work on a project that both are involved in. Capitalization takes reuse a step further in order to make sure the costs incurred from an initial AI project are shared across other projects.

The notion of reuse and capitalization can be seen through the following example. If a company is working on four primary use cases to jumpstart AI efforts, the organization can also work on other smaller use cases by reusing various parts of the four main ones, therefore eliminating the need to start from scratch with data cleaning and prep, operationalization, monitoring, and so on. As an added bonus, this approach can also help teams uncover hidden use cases that can drive more value than originally anticipated, opening up new pockets of potential profit or cost savings. To learn more about how to efficiently leverage Enterprise AI, check out the full white paper, ***The Economics of AI: How to Shift AI From Cost to Revenue Center***.



Reuse and Capitalization at Work

In the pharmaceutical space, for example, organizations can build a model for a specific disease in order to predict when a patient is ready for a specific kind of drug therapy.

Once the first model is built, it can be scaled and repurposed for treatment of another disease using the same drug therapy as the first disease, enabling teams to avoid having to start from scratch, reducing the time to model completion, and injecting efficiency throughout the process.

Before we give specific examples of how Dataiku is useful for optimizing the TCO for a data science platform, it is important to remember that each organization is in a different stage of their journey to attaining this organizational, data-driven culture that transcends teams, processes, and technology. Some organizations may have fundamental use cases in place that allow the business to optimize costs and accelerate its ability to execute on mission-critical business functions, while others may not be in this position.

Reducing the TCO for a Data Science Platform With Dataiku

When organizations are running a TCO to determine the optimal data science and machine learning platform for their organization's needs (and look beyond the value that is being promised), there are a few things to keep in mind.

First, it is important to compare apples to apples. Are the solutions comparable from a capabilities perspective? Next, the costs involved with each platform should be properly scoped. What should be included? What existing costs will be impacted by the platform? What costs will remain the same? What are the implementation and running costs of the new platform? What are the staff costs involved? Does one platform offer a beneficial impact on cost versus another? Are there any hidden costs? Finally, run different customer scenarios, taking into account team size, data needs, and complexity levels of varying analytics projects.

Below is a non-exhaustive round-up of how we believe Dataiku can drastically reduce the TCO for data science initiatives. The end-to-end platform:

1. **Enables data prep and *AutoML***, removing the cost of licensing, enablement, and integration for separate data prep and AutoML products
2. **Is completely server-based**, removing the cost of maintenance of desktop-based products
3. **Enables the management of Spark clusters**, removing the overhead of paying for a Spark management solution
4. **Comes with a strong operationalization framework**, removing the need to build a fully-fledged CI/CD code-based framework (on top of the existing solution)
5. **Provides a clear upgrade path** with no need to transition platforms or migrate in the future
6. **Minimizes the overall number of tools** by offering one truly comprehensive platform, avoiding the need to cobble together multiple tools for ETL, model building, operationalization, and so on
7. **Promotes reuse** via capitalization, allowing the organization to share the cost incurred from an initial AI project across other projects, resulting in multiple use cases for the price of one, so to speak
8. **Is future-proofed and technologically relevant**, helping teams avoid significant upgrade costs or lock-in when faced with limited infrastructure options that can hinder growth

Dataiku manages the entire data science pipeline for teams in a way that is flexible, collaborative, and governable. With every use case that is added to the mix and every project deployed, organizations are creating more **"AI debt"** that must be serviced, consequently decreasing team productivity (a problem that is especially exacerbated without a clearly defined MLOps strategy). If teams need to start from scratch with each project, eventually AI productivity grinds to a halt.

Now is the time for organizations to reassess their AI use cases in order to maximize ROI and maintain new value, making sure that reuse and capitalization is cornerstone across each use case. By finding ways to generate efficiency gains and cost optimizations, organizations will be able to leverage data science and AI as the gateway to becoming a smarter organization.

The Role of MLOps in the Next Wave of AI

MLOps — the standardization and streamlining of machine learning lifecycle management — will undoubtedly play an ever-increasing role in the future of AI and how organizations continue to reach new stages of their AI maturity. Business needs shift and results need to be regularly relayed back to the business to ensure that the reality of the model in production (and on production data) lines up with expectations and still addresses the original problem or meets the original objective.

A sound MLOps practice will be a critical component to scaling machine learning efforts, i.e., going from one or a handful of models in production to tens, hundreds, or even thousands that have a positive business impact. MLOps best practices will enable teams to:

- 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 have more optimal performance).
- Ensure that model performance is not degrading in production.

Further, MLOps helps provide insulation from risk, as it can become tricky for teams to maintain a global view of the state of each operational model without some standardization. Pushing models into production is just the beginning of the performance monitoring phase — teams need to make sure the model acts as expected, adjusting when necessary, in order to truly mitigate potential, tangible risks that can be detrimental for business.

> ***"Data and analytics have become core to how organizations serve their customers and optimize business processes. They are the foundation of new transformational business models, revenue streams, and process and cost optimization."***
>
> *- Gartner, Top 10 Trends in Data and Analytics 2020*[5]

[5] Gartner "Top 10 Trends in Data and Analytics, 2020," Rita Sallam, et al, 11 May 2020. https://www.gartner.com/document/3984974 (Gartner subscription required)

Conclusion

By taking a precise approach to value extraction at each stage of AI maturity, organizations will be able to implement smarter business processes, garner an improved technology stack and team efficiency, harness agility to accelerate output and time to value, and enhance risk mitigation and compliance. With ever-more volatility and complexity underscored by the global health crisis, there's a growing need for resiliency and value generation.

Ultimately, the goal is a broad and inclusive organization in which everyone is working toward solving business challenges from the same data. While cost reduction refers to just that, driving down costs associated with AI applications, it also means creating consistent value with AI and improving decision-making capabilities.

For organizations focused on getting to the next wave of AI (i.e., from Establish to Expand), here are some actionable steps to take now:

- **Aim to break down data silos** so that data projects are no longer only limited to "experts"
- **Consider tool consolidation** so teams don't need to transfer between a myriad of tools and can do everything — from data prep to production — in one place
- **Begin exploring MLOps** to streamline the process of maintaining models in production, reducing delays for new deployments
- **Evangelize reuse and capitalization on previous projects** to avoid significant rework and save costs (both from a bandwidth and technology perspective)

"Companies should begin now to game out the potential impact of pervasive intelligence on their business and their industry to position themselves to reap the benefits."

- Deloitte[6]

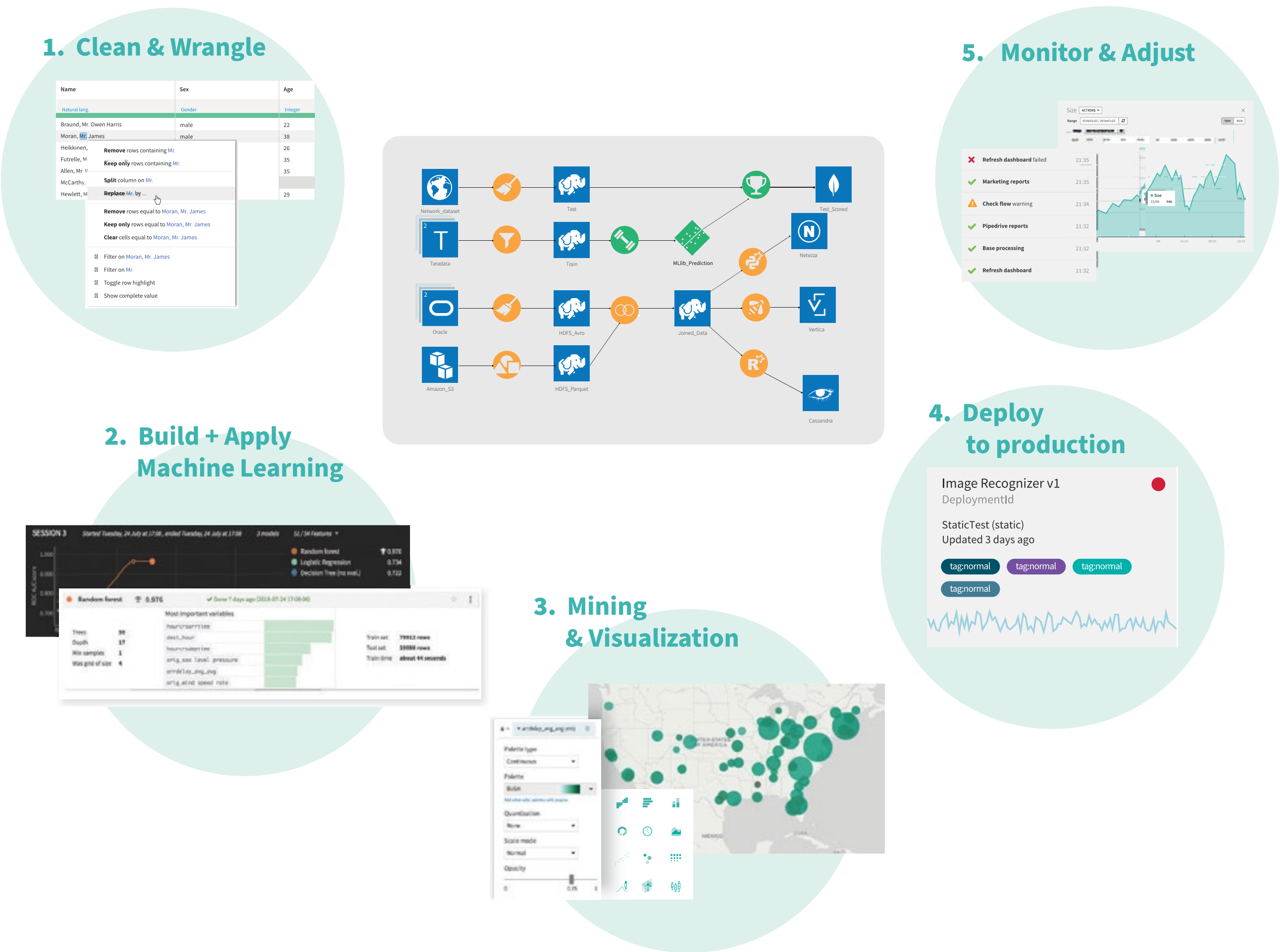
[6] https://venturebeat.com/2018/11/08/deloitte-pervasive-ai-promises-to-transform-agriculture-health-care-and-manufacturing/

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



data
iku