research

ALGORITHMIA

Building versus buying an ML management platform

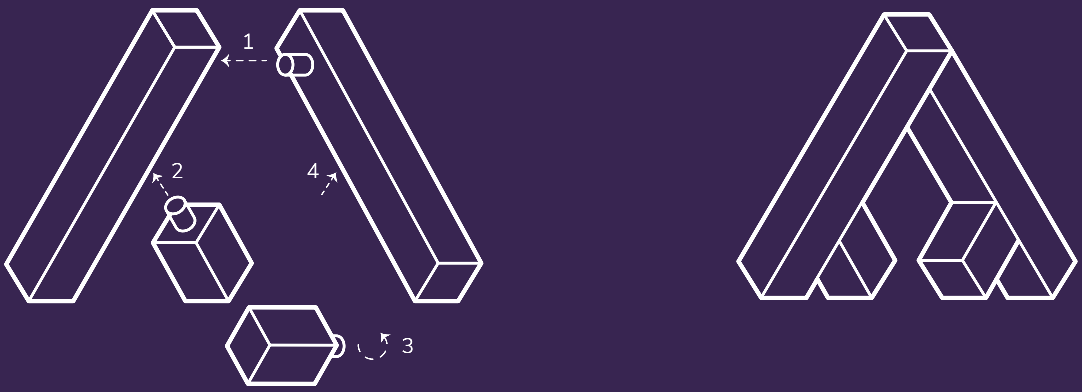


fig. 1 build

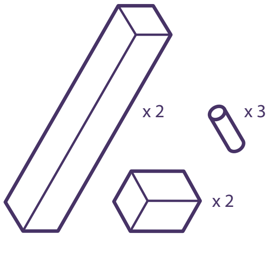
fig. 2 buy

Abstract

Determining whether to build or buy a machine learning operations and management platform will drastically impact your company's competitive standing in its industry. Time wasted up front will not be regained down the road, so extracting value from ML as soon as possible to maintain a competitive advantage in your industry is paramount.

This whitepaper will discuss the following sections to help you make an informed decision for your organization's AI future.

1. What building a machine learning management platform entails.
2. How to make a business case for ML and share it within your organization.
3. What you should evaluate when looking to buy an existing ML management platform.
4. Our assessment that purchasing an off-the-shelf platform is the best solution.



Introduction

Forecasts estimate that AI will add more than $14 trillion USD to the global economy in the next 10 years despite the recent global economic disruption (Wall Street Journal). Ample opportunities exist to capitalize on that within every industry today. In fact, 48 percent of industry-leading firms now compete with algorithms compared with 5 percent a decade ago (NewVantage Partners). By next year, that number will be over 50 percent.

The speed at which machine learning is transforming entire industries means your company cannot afford to waste time on tooling and infrastructure instead of deriving insights from an ML portfolio now. You need a machine learning operations plan now. Making AI–minded decisions starts with this question: **should I build or buy a machine learning management platform to operate my ML lifecycle?**

48%

of industry-leading firms now compete with algorithms.

According to Gartner, 85 percent of all AI projects fail, and the majority of organizations actively developing a machine learning capability are struggling to extract a return on their AI investment (Algorithmia). Therefore it is crucial to know up front what to expect in terms of infrastructural requirements, developer workloads, time, and costs associated with building an in-house machine learning management platform so you can prepare to meet the goals you want.

Our assessment is that in order to extract value as soon as possible from AI and maintain a competitive advantage in your industry, purchasing an off-the-shelf platform that fits into your existing workflow is the best answer.

Let's start by identifying general challenges in machine learning and how to position your company to best overcome them.

The challenges of machine learning

Without the knowledge, experience, and skills needed for ML operations (MLOps), it is not easy to build and manage complex ML infrastructure or to make crucial decisions on how to deploy models quickly, scale them efficiently, and secure them without incurring excessive costs. They may also struggle to find experienced talent, manage infrastructure configuration complexity, and overcome ML–specific tasks, such as autoscaling and versioning.

And technologically speaking, the largest barriers will be managing multiple frameworks and following expensive and time-consuming authentication and security protocols.

Let’s take a more in-depth look at some of the challenges:

Labor/talent

Getting machine learning up and running is labor-intensive, requiring dedicated staff across several job functions. If a company does not have the headcount, it must hire for those scarce resources, which can take time and be costly.

Processes

MLOps and management are not like regular software operations and management that DevOps engineers know. Data scientists may understand machine learning and how to train, build, and test models, but they may lack experience in ML operations or model deployment across an organization. So too, DevOps engineers rely on specific infrastructure and integration practices to work but may lack understanding of niche ML concepts like model drift and accuracy changes.

Tooling inflexibility

Integration among multiple systems is key for running data through your models and for more complex functions like pipelining and versioning. Companies wanting to develop ML business cases will need the latest, state-of-the-art ML technologies, and there are more and more in development all the time, meaning your infrastructure must evolve over time in several ways:

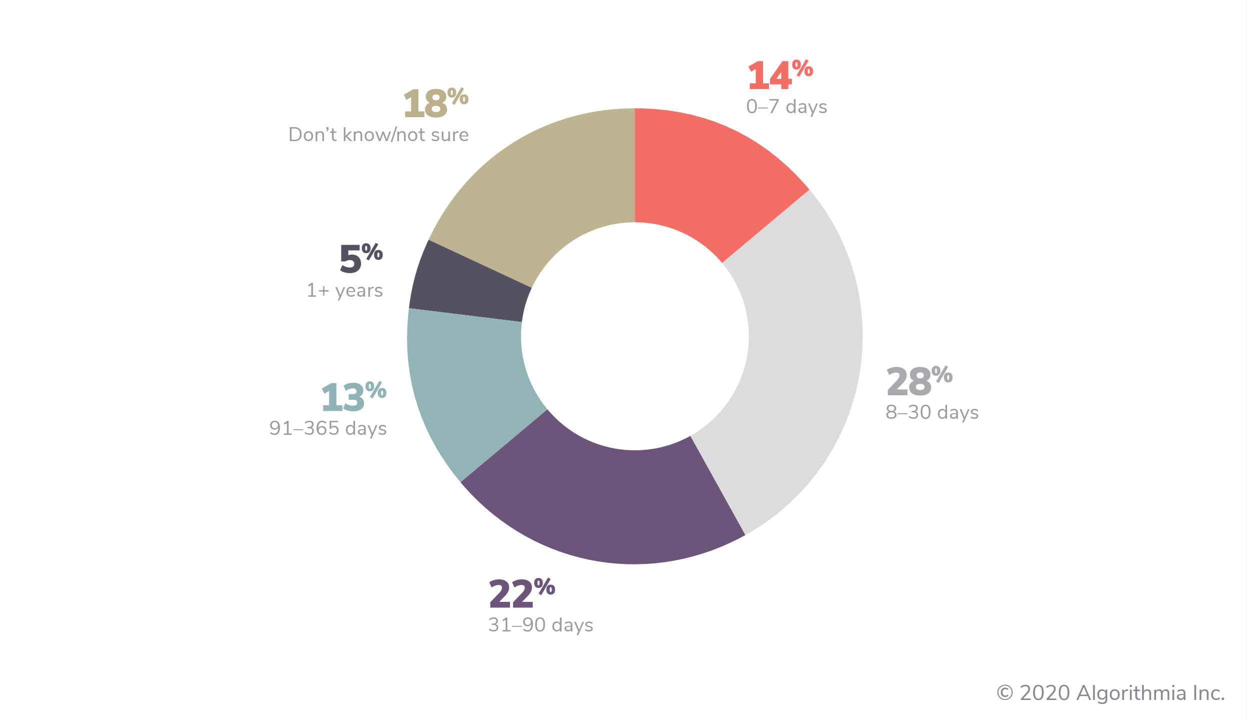
- Data connections: prepare data sources and pipelines
- Model deployment: load, catalog, version, validate models
- Operations: manage cloud costs, continuous integration and deployment
- Governance: evaluate model performance, maintain ML portfolio health
- Security: data, models, infrastructure, and security permissions

Once these challenges are solved, there still exists one of the biggest obstacles to achieving value from machine learning at the enterprise level: getting models to production.

The long road to production

Half of companies say they spend between 8 and 90 days deploying one model (Algorithmia). How long will it take your organization to deploy and serve a single machine learning model? How about 10, 20, or hundreds of models?

Model deployment timeline



© 2020 Algorithmia Inc.

Organizations that do not have a dedicated machine learning operations and management platform end up wasting ample data scientist time on operationalizing models. And if an organization is struggling to get value from its ML projects, it could be because its AI talent is consumed with infrastructural tasks instead of the higher value work of applying AI/ML to drive business insights and decision making. Those infrastructural tasks can be done in a few minutes with a robust machine learning operations and management platform.

Market conditions, models, and data change rapidly. To maximize windows of opportunity, you need to quickly and efficiently put models to work. ML–based applications waiting in a DevOps queue jeopardize the value they can deliver. To that same end, a prediction that comes even one day too late no longer provides a competitive advantage.

Machine learning deployment platforms

In response to widespread enterprise AI deployment pains and mounting pressure to deliver on AI promises, the machine learning operations and management market emerged. The global ML market was valued at $1.58 billion in 2017 and is expected to reach $20.83 billion by 2024. With this demand for applied ML, most organizations will need to move quickly to remain competitive.

Building a machine learning management platform

> Given the current economic uncertainty, the state of hiring, and the demand for cloud talent, securing technical resources is the biggest challenge to overcome in a build scenario.

The effort to build machine learning operational infrastructure internally is often underestimated by business leaders and their developer teams who build and maintain software platforms as part of their day to day.

To build a minimum viable offering to deploy, serve, and manage machine learning models at scale, consider team resourcing, development components, scope, cost, and timing. Even if all these assets converge, it is still likely that a do-it-yourself solution will be far inferior in comparison to available market solutions that have been developed over a much longer period of time with many more resources.

Let's begin by laying out some questions you should answer before proceeding with plans to build a platform internally:

- Can you afford the delay in building time to offset the licensing costs of buying a solution? What happens if your competition delivers AI faster than you?
- Can your team commit to full-time machine learning deployment, platform development, maintenance, and technical support staff over the entire lifespan of your AI program?
- Is development and maintenance of an existing solution the optimal use of your organization's time and budget?
- Can you efficiently deploy machine learning models at the speed of business? How long will it take you to deploy one model? Many models? How will you version models?
- How many resources will you need to develop and manage the system?
- Can your team monitor and update deployed models continuously? How?
- How will you maintain the security of your data, models, and the infrastructure that they run on?

Below, we'll go into detail about how these questions should be considered in a build or buy decision. In our comparison, we separate initial development costs from ongoing support and maintenance. Given the current economic uncertainty, the state of hiring, and the demand for cloud talent, securing technical resources will most likely be the biggest challenge to overcome in a build scenario.

Plan the project

After securing sponsorship, it is time to start planning your project. To build, you'll need to establish a cross-functional project team, start thinking through the initial project requirements, scope the level of effort needed, develop reasonable timelines, and mitigate a number of risks up front. You will also need to be realistic in your estimations of resources needed. At a high level, you will need 18 technical resources, 28 months of time, and approximately $2.8 million to build a viable machine learning management solution on your own.

> At a high level, you will need 18 technical resources, 28 months of time, and approximately $2.8 million to build a viable machine learning management solution.
> Tryolabs, 2019

It is very likely that within that timeframe, you will complete agile milestones, such as a minimum viable product after six months that can do source control but little else. Then, a couple of sprints later, you'll introduce the next feature set on your roadmap. Bit by bit, you'll build components of an MLOps platform, relying on open-source software that you'll have to customize for your use case, but it will be piecemeal and your ML goals will be on standby in the meantime.

Put together the team

Your build project team should include stakeholders and contributors from both the business and IT sides of your company as well as compliance, data governance, enterprise architecture, data engineering, data science, security, and support. The ideal team includes the following resources:

- *Executive sponsor*
- *Project manager*
- *Solutions architect*
- *Senior cloud developer*
- *Full stack developer*
- *Cloud architect*
- *DevOps engineer*

Timeline to build a machine learning management platform in-house

Phase	Time (in months)																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Architect and design	▨	▨																										
Build execution layer			▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨															
Build, deploy, and serve													▨	▨	▨	▨	▨	▨	▨									
Model sharing and discoverability																			▨	▨	▨	▨	▨	▨	▨	▨		
Authentication, RBAC, and security																									▨	▨	▨	▨

The following six-phase project plan provides a high-level overview of a 28-month minimum viable product build. This is based on a scenario in which the personnel prioritize building an enterprise-ready ML platform. Note that in our experience, over the course of 28 months, resources will be pulled into other priorities, and dedicated staff will be expected to handle multiple concurrent projects and tasks outside the scope of this project.

As such, it is highly likely that a build project will require more workers, time, and resources than our conservative estimates to become enterprise-grade. Organizations should consider that additional risk to their return on investment and value extraction from such a massive and prolonged resource expenditure.

Ensure guaranteed value from your investment

"Most ROI will be seen in cost reduction and efficiency, as that's how AI is currently used. However, as enterprises evolve their AI expectations and projects, the technology will mature to have more transformative and strategic impacts" (Gartner).

Up-front investment for DIY ML management platforms will significantly impact the time to value and return. Long and inefficient ML lifecycles (about 18 months to deploy the first model in a DIY scenario) mean you cannot fail fast, learn from iterations, or capture business opportunities in time.

Gartner warns that CIOs should "ensure the desire to push forward with a popular technology doesn't overrule realistic drawbacks and planning. The hype itself can be a problem, alongside other logistical and strategic challenges." Doing AI for AI's sake will likely not yield much business value; but deciding to outsource an ML platform solution will increase the likelihood that your organization will not just achieve ROI but also deliver business value above the top line sooner.

Six-phase plan

Architect and design; months 1-2

A robust planning phase takes an average of two months before work on the project itself can commence (Algorithmia). During this phase, several cloud architects and a project manager gather requirements to determine what is needed for the company's specific build. They will determine scope, risks, resources, and most importantly, they design measurable business use cases. This team will also evaluate needed technologies, hardware, and software services during this phase.

Additional requirements include:

- Deciding which programming languages and machine learning frameworks your platform will support
- Defining API interfaces to existing company data
- Determining which services to use for the execution layer and where to host container images
- Collecting required security and compliance requirements
- Gathering cost and time estimates for software
- Producing a project plan and requirement documentation to pitch to senior stakeholders

Build the execution layer; months 3-13

During the next 10-month phase, you will need multiple senior developers, at least one cloud architect, and a project manager. This group will determine which software packages data scientists need, build a library of machine images and docker container images, and automate image creation so it can be done repeatedly.

The container runtime execution layer will be designed to serve APIs, support unit and integration tests, scale on a per-model and use cases basis, ensure efficient hardware and framework options for CPU/GPUs, Nvidia drivers, CUDA, TensorFlow, PyTorch, and other popular data science frameworks.

For model serving, the team will also need to design and build a monitoring and recovery solution for the platform that contains appropriate alerting. The system will need to be able to do the following:

Facets of efficient model serving
Elegantly autoscale
Recover from failures
Retry API calls
Save detailed logs for troubleshooting
Minimally provide reports for deployments, usage, and system health

Additional requirements include:

- Updating and maintaining runtime languages and frameworks
- Integrating and connecting to external and internal dependency management solutions
- Scaling container caching and data caching
- Optimizing infrastructure scale-down
- Updating and maintaining scheduling resource framework

Build, deploy, and serve; months 13-19

In this six-month stage, you will need several senior developers and a project manager. They will transition the project code into a running service with source control integration. At this point in the project, they will also create a container repository for system images with appropriate security and access controls configured.

To enable crucial model deployment version control capabilities, the team will also develop needed functionality to deploy and monitor new machine learning models with no downtime or rollback and will upgrade models without breaking changes for downstream systems.

Additional requirements include:

- Standardized REST APIs for integration
- Method of scaling underlying infrastructure
- API key creation and versioning system
- Data API security system

Model sharing and discoverability; months 19-26

In the model sharing phase, you'll need full-stack developers, at least one cloud architect, and a project manager. Throughout this seven-month stage, your team will design and create an easy-to-use UX, a browsable catalog of machine learning models, and ensure proper authentication/integration with the rest of the system components.

Additional requirements include:

- APIs to push and manage the platform to existing pipelines
- Algorithm promotion/demotion system
- Model tracking and evaluation toolset

Authentication, role-based access control (RBAC), and security; months 25-28

To wrap up the basic build of a minimum viable offering, the last construction phase should be spent developing robust authentication and security measures according to the organization's security and data access standards.

Typically this phase of the project takes at least three months for several developers to build. Common baseline functional security requirements include but are not limited to user sign in, role-based feature access, integration with OAuth or other enterprise authentication mechanisms, and data access security.

Additional requirements include:

- Container auditability
- Chargebacks/consumption reports and usage metrics
- Traceability and auditing
- Model logging

On-going maintenance and support

After the team rolls out the finished product, the machine learning lifecycle is set in motion and will need to be continuously maintained. At least two DevOps engineers and four developers will share the maintenance and support work. This is also the period of time to offer training to departmental users of the system.

For estimating long-term maintenance work, be sure to plan how you will respond to incidents and make frequent updates and fixes to machine learning frameworks, machine learning and cloud libraries, hardware, model performance, and scaling resources.

Remember, this is a production system. Your team will most likely need to ensure 24/7 support. Minimally, plan for quarterly updates to base software installations, updates to machine images, docker images, and out-of-band updates for critical software vulnerabilities.

Additional requirements include:

- Handling and testing failover for HA/DR
- Ongoing personnel staffing
- Organizational/regulatory security changes

A healthy, functioning machine learning lifecycle must be tuned continuously to deliver business value. So though, there is a heavy lift up front to build the infrastructure, models, and support mechanisms to achieve a minimum viable platform, it is by no means the end of the project. Dedicated workers will need to maintain the program's systems thereafter.

Let's turn now to options for off-the-shelf solutions.

Buying a machine learning management platform

There are many reasons why enterprises buy proven machine learning management platforms instead of attempting to build them. A solution that makes good business sense accelerates getting your models into production, provides consistent service quality, and includes predictable operational costs among many others.

Benefits of an off-the-shelf ML management platform	
Protects organizations from expensive cloud-host lock-in	**Prevents planning delays**
Allows developers and data scientists to focus on true areas of expertise	**Repurposes existing models in production for new, more complex operations**
Reduces development, maintenance, and technical support requirements	**Increases speed of innovation by removing day-to-day reliance on engineering teams**
Amplifies machine learning ROI by supporting more models in production without additional labor costs	**Improves data democratization across an organization with broad language and framework support**
Leverages existing IT operations talent with more enterprise-friendly, efficient, and secure approaches to deployment	**Reduces security and data privacy risks**
Includes autoscaling support	**Minimizes compliance risks**
Provides a consolidated view of an ML program's health and impact	**Ensures accountability to keep stakeholders happy**
Eases operationalization friction	**Provides predictable, ongoing operations and management costs (ie. OpEx)**
Reduces time to deployment for models	

In contrast to a build project, buy solutions require fewer upfront resources, reduce the burden or fixes and tasks for your team, and drastically reduce cost and time to complete. You also get a proven enterprise-ready solution instead of a minimum viable build solution.

The predictable and incremental costs of subscribing to an ML operations platform from a vendor enable budgetary planning in a way that building a platform does not. Third-party options often support flexibility in speed, ease of deploying many models, and have the capability to version those models in support of business applications and customer datasets.

The buy route also begins with questions that should be answered before making a purchase:

- Is the solution easy to use/understand? Will it reduce risks?
- Does it integrate with my organization's current workflow and tools?
- Will it cut down on team infrastructure tasks and/or create more time?
- Will it enable faster delivery of ML value to the business?
- Does it allow teams to collaborate more efficiently (ie. is it extensible across the organization?)
- Will it help us save on cloud spend?
- What types of reporting/alerting does it have?
- Does it align with enterprise technology standards, such as security and compliance requirements?
- What value will this tool ultimately provide (ie. reduce cost, cut down on staffing resources, automate workflows, etc.)?
- Does the tool support our current skill sets and/or provide a capability we do not have or would take too long to develop?
- Is it an effective, cutting-edge solution?
- Is it cost effective and within budget?
- Is the time it takes to learn the tool worth the investment (ie. is the ease of adoption acceptable)?
- Does this tool future-proof workflows and prevent lock-in?

Your research into off-the-shelf solution providers should yield answers to these questions. Vendor solutions that do not address these questions likely have capability gaps that will arise during onboarding and should be excluded from your selection criteria.

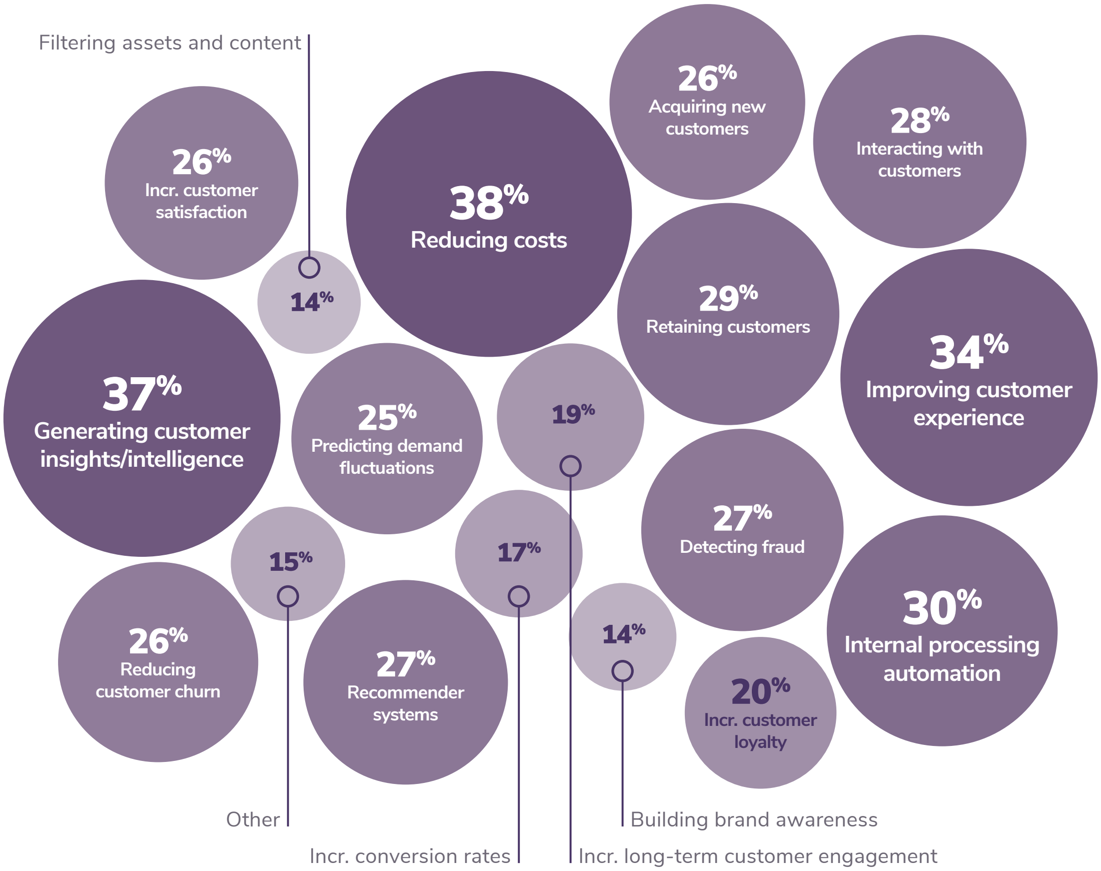
Next let's explore what a buy scenario looks like.

Make a plan

To begin your buying process, you'll need to align organizational stakeholders toward a common machine learning outcome. Determine what is the most important outcome(s) to achieve. Is it reducing personnel or software costs? Gathering customer insights? Improving customer experience? In our 2020 state of enterprise machine learning report, we asked organizations about their ML use cases.

Reducing costs, generating customer insights, and improving the customer experience ranked as the most common machine learning uses sought by enterprise companies, however, there are a myriad of opportunities to leverage ML for business value.

Machine learning use case frequency



Build a team

The personnel involved in a buy scenario are fewer than for a building but are more representative of the teams that will be *using* the solution and as such, must understand the specific needs and intended purposes.

The team will include stakeholders from across the organizations, including but not limited to the resources shown below:

- *Executive sponsor – budget discussion*
- *Project manager – requirements checklist*
- *Cloud architect – deployment checklist*
- *DevOps engineer – connect to APM platforms*

Vetting criteria

There are several factors to consider in a buy decision for a machine learning management platform. Without careful planning, a company might unwittingly lock itself into a needlessly expensive or resource-intensive contract, determining too late that a new framework is not compatible with it or that the platform has poor capacity-provisioning capabilities.

> Look for an ML management solution that allows for flexible configuration.

Our advice: look for an ML management solution that allows for flexible configuration. Any company with unique ML use cases—likely all companies—or proprietary datasets will need built-in flexibility to anticipate changes in tooling, end goals, and available resourcing.

The criteria for determining platform viability and suitability are as follows:

Fit: how well the solution handles workloads similar to yours. Can the provider demonstrate its ability to handle your specific use cases? What about customer-facing applications, internal optimization and analysis applications? Application platforms, including web, mobile, IOT, desktop, legacy, and batch processing?

Users: who the solution is fit to serve. Verify that a platform fits with your company and intended personnel. If it's more conducive to a team setting with a user limit, but you need a platform that can be accessed by multiple teams across your organization, perhaps in different countries, at any time, that is not a good fit.

Applications: determine what the platform is designed to handle. Is it built for cloud scale? What about continuous mission-critical applications? Consider tomorrow's needs in the vetting process, as in what might the company need in 2-5 years as opposed to what it needs now.

Hosted: where is the platform hosted? This plays into the next category—security—as it may have compliance and regulation implications, so ensure the platform is able to fit your company's requirements. If you want to run models near where your data is stored, you need to find a platform that can provide an on-premises deployment.

Security: consider any potential platform's encryption options and your company's security requirements. Your data will run through the ML platform; consider the implications of that when vetting solutions. Also consider that ML model security and ML infrastructure security are two distinct types of governance you'll want to ensure are covered. Discuss the risks with each solution provider and how they mitigate them.

Support: Having a dedicated on-call support feature may be the most important part of the determination process, especially if you're planning on mission-critical ML applications. Do the options you're looking at provide 24/7 support, and how easy is it to communicate with their team?

Billing: What does the billing/invoicing process look like for each platform candidate you're vetting? How do they charge and does that model fit your ML workflow needs? Are consumption-based billing models available? Are there managed service options?

Keep these criteria at the forefront of your buyer process and don't compromise on any one of them.

Make your case

CIOs and other technical executives make investment decisions with a bottom line in mind. To make your case in this build v buy scenario, it's important to put together a cost-benefit analysis. Your review of options should include both business needs and justified costs supported by an estimated return on investment ROI, payback period, and risk assessment.

> “If your company is not currently ML–minded, rest assured your competitors are, and the rate of AI's development is bound to increase exponentially. Now is the time to future-proof your organization with AI/ML”
> 2020 State of Enterprise Machine Learning report

Determine the cost of doing nothing

The impact on profitability and ROI drives most business decisions. Whether your organization is thinking in the short- or long-term, expected business impact is what dictates which plans get implemented and which plans fall by the wayside. The cost of not implementing machine learning (neither building nor buying a machine learning deployment platform) is detrimental to your company's future.

Examine your organization's current model development processes and forecast what they will look like when your organization needs to manage hundreds or thousands of machine learning models. By working through that exercise, you will see specific opportunities and cost trade-offs to consider in your ROI estimates.

For example, if your organization continues down the path of disparate departmental software deployments, then you'll likely end up with siloed or duplicated programming efforts. It's imperative to identify those risks and whichever solution you choose will need to solve it to deliver value to your organization.

Challenges of building an ML management platform

Build	Buy
Compounding data collection expenses	Elastic inference for cost management
Personnel resources	Generate business returns quickly and repeatedly
Complex, time-consuming ML application planning	Scale across the organization automatically
Manual deployment for each new model	Efficient model operations from a central model management system
Cloud service fees	No planning delays
Fluctuating SLAs relative to business requirements	

The value of machine learning in production

Earlier, we discussed types of value to get from a built machine learning platform. The buy scenario contains several types of value as well:

1. Time savings value
2. Leverage and risk value
3. Spend/cost savings efficiency

Time savings concerns time to value—typical model deployment times for Algorithmia customers are under 6 weeks (compared with an average of 12-18 months per model in a do-it-yourself scenario).

Leverage and risk value is the ability to generate business returns and take market share with rapid deployment of ML in volume and agility. It's iterating and capturing shorter windows of opportunity, and leveraging learnings, infrastructure, and existing talent.

The need for specialized engineering experience when building an ML management system is greatly reduced when you purchase an MLOps platform, including security and governance requirements.

Cost savings efficiency means predictable cloud service and consistent SLAs. Cloud consumption costs with Algorithmia are 2.5 to 7 times less than do-it-yourself cloud-managed consumption, according to Tryolabs, an independent third-party benchmarking report.

What is your ultimate goal for machine learning?

As shown, there are a number of factors that surround the decision to incorporate machine learning into your company operations. Many of them will be company-specific so ensuring you have a clear picture of what you want to gain and your means of gaining it is paramount.

Use cases for ML vary from company to company and across industries and will play into what kind of platform your company will need. But what will remain constant across all industries is the ML timeline.

At the end of the day, you need to determine how to get the shortest time to value for machine learning while reducing the risk of failure or stalled efforts. Building an internal platform will be the longest route by far and may not deliver on value add. A flexible model management platform designed by an experienced team and built for evolving complexity vastly reduces risk and ensures your company can extract the most value from your machine learning models.

About Algorithmia

Algorithmia is a leader in the machine learning model deployment and operations space. We empower every organization to achieve its full potential through the use of artificial intelligence and machine learning. We provide the shortest path to AI value by delivering model deployment at scale for enterprise workloads.

More than 100,000 developers, Fortune 100 companies, government intelligence agencies, and private organizations trust Algorithmia to:

- Deploy models from most frameworks, languages, or tools
- Connect to existing data sources or create new connectors
- Scale model inference on cloud or on-premises infrastructure with uptime SLA
- Manage your machine learning lifecycle in a simple model management system

To learn more about how Algorithmia can help your company deliver AI value sooner,
visit our website at algorithmia.com.

ALGORITHMIA