TECH BRIEF

# How to get started with Machine Learning: Expert advice from the front lines

Smart companies know that artificial intelligence and machine learning have the potential to transform business, but many aren't sure where to start. Here's an expert guide based on 10 of the most burning questions asked about what it takes to launch a machine learning initiative.

**Larry Pizette**
Senior Manager, Machine Learning Solutions Lab
Amazon Web Services

## INTRODUCTION

Machine learning (ML) is frequently the catalyst that turns business data into accurate predictions and actionable information. For many companies, however, the barriers to entry seem daunting:

- Data needs to be discovered, organized, and standardized
- Compute resources need to be sourced to handle ML workloads
- Models need to be built and trained by data scientists
- Outputs need to be integrated into business processes

When faced with requirements that are hard to scope out and goals that require imagination, companies need expert help. Larry Pizette, senior manager and leader of the Machine Learning (ML) Solution Lab at Amazon Web Services, knows a lot about overcoming ML barriers. He meets with customers from diverse industries, helping them plan and implement their ML initiatives across a wide range of use cases.

We had the opportunity to sit down with Larry and pose 10 questions, eliciting his insights for business leaders who want to understand ML trends, how to get a project started, and how to move from proof-of-concept to production.

## APPROACHING ML

### Q1 What's driving interest in machine learning today?

We're at an inflection point, where machine learning has hit the common consciousness. There are references to artificial intelligence and machine learning on television, magazines, websites, and product announcements – it's everywhere. Customers are looking at ML for a variety of different reasons, but it tends to be around the same goals: better experiences and outcomes for their customers, improved operations, and, for commercial customers, better financial results. Business owners want to unlock the potential of their data. It's hard not to pay attention to machine learning.

### Q2 Are there specific industries or functions that are strong candidates for ML?

We have customers from many different business areas: oil and gas, healthcare, life science, insurance, financial services, restaurants, professional sports, transportation, government, and educational institutions. Customers are coming to the ML Solutions Lab because they want to get started with ML capabilities in their organizations, and they'd like to understand the art of the possible.

AWS TECH

### Plug-and-Play ML

Many leaders think ML is complex and labor-intensive, but it need not be. AWS can get you started today, with no ML expertise required, through a range of ML services that can be easily added to your applications:

- Analyze your images and videos with Amazon Rekognition.
- Build voice and text chat bots for your business with Amazon Lex, which is powered by the same technology as Amazon Alexa.
- Create applications that talk by turning text into lifelike speech with Amazon Polly.
- Discover relationships and insights in text, such as your customer service emails or transcripts, with Amazon Comprehend.
- Localize your content, such as websites and applications, quickly and cost-effectively with Amazon Translate.

"Getting started is really a matter of three steps: learning about ML, ideation on use cases and approaches for validating the ideas, and implementing one or two high-value proof of concepts."

For example, insurance companies want to help assess risk better, to better price their products and maintain their profitability. Business leaders, government leaders, educational leaders are all starting to think about how ML can impact their organizations. They know that they need to get started; they're just not 100 percent sure where to begin. They know that Amazon has an extensive history with machine learning, so they come to us, and we can help them get going.

## IMPLEMENTING ML

### Q3 | What steps do organizations need to take when they want to get started with ML?

Getting started is really a matter of three steps: learning about ML, ideation on use cases and approaches for validating the ideas, and implementing one or two high-value proof of concepts (POCs).

If your organization doesn't already have a solid foundation of ML knowledge, we recommend taking the time to educate your business and operational leaders. When customers come to us needing this knowledge, we first engage them in a machine learning discovery session, sitting down with our specially trained business development organization. As part of that step, we make sure that customers understand, at a business level, how ML works, and the importance of data to their organizations.

The next step is typically to have an ideation session and identify POCs. We like to include both business leaders — the business owners that are looking for the outcomes that they want — and IT folks that can help us identify the data to meet the business need. This is the point where we start identifying the details of a POC and the business outcomes.

We also dive into the importance of having an understanding of the data, and in many cases, having annotated data — that is, having data that is tagged with information about the data itself. An example would be an image and a description of the image. This step is significant, as many organizations don't know the data that they have or how they'll access it — and data is necessary for training ML models.

## PROVING ML'S WORTH

### Q4 | So that brings us to the third step. What does a successful proof-of-concept look like?

We find that customers that are willing to identify impactful business cases — something important that can really make a difference to their organization — are the ones that can demonstrate results quickly with ML.

It's really important that customers are willing to prototype and do proofs of concept to get going. If they try to "boil the ocean" and have a multi-year deliverable, it's much harder to demonstrate forward progress.

Where I've seen POCs go really fast is when customers have the data and are highly motivated. With those customers, they get the data; they allocate resources to the collaboration with us; they get us going with access to the data and a clear objective; and we can show initial POC modeling results in as little as three weeks from when we get access to the data.

### Q5 | Once the POC stage is finished, how long does it take to implement an ML solution in production?

Once you've proven the potential of ML, the next step is to move capability to production, which may include integrating the ML capability into a larger IT system.

When a customer moves a workload to production, it's typically longer in duration

### Is Your Data ML-Ready?

Data is gold for leaders who are looking to disrupt their industries with ML. The question is, is your data ready? Many companies only begin working on data organization as they plan a proof-of-concept, but you need not wait. Follow the guide below to put your organization in a stronger position to implement and scale ML.

**1. Locate your data**
Go on a journey of discovery to uncover all of the data sources in your business. Explore different functions, divisions, systems, and processes to capture and categorize the different types of data you possess. Everything counts, from structured data like billing and CRM to unstructured data like web logs and social media feeds.

**2. Evaluate your data**
Only quality data will deliver the ML benefits you want. Before proceeding too far, it's important to determine how much of your data is actually usable. Databases are often incomplete, differ in organization or taxonomy, and/or are duplicative with regard to content. Taking the time to assess the data and identify the fixes now will save considerable time and expense down the road.

**3. Scrub your data**
Now it's time to get the data into shape so it can be used to power your ML goals. Cull duplicate information, fill in missing fields or details, correct any errors, and get it organized. You will no doubt have data that cannot be neatly normalized to comply with models or workflows. Just flag and separate it to prevent it from undermining development of your algorithms.

**4. Centralize your data**
Maintain and protect your new, high-quality data by storing and managing it centrally. This is also the time to add the annotation or tagging needed for machine learning systems to accurately analyze the data and transform it from describing what's happened in the past, into powerful predictions and action for your business.

"We recommend focusing on how you can deliver a better experience for your customers and identifying the business and operational outcomes desired — and then leveraging ML to meet those needs."

than the POC. It really depends on the complexity, because we are building machine learning into the overall system. If it's a standalone type of capability, it can go into production very quickly. If it has to be integrated with a much larger ecosystem of IT systems and business processes, it can take several months.

But the cool thing about it is, once you get up and running, you can keep updating your models. You're gathering data, you're gathering outcomes, and you're improving your models over time. As you gather more data and information on what "good" looks like, it's critical to iterate by re-training and re-evaluating your models — this will help you to ensure you're getting the best outcomes possible.

As there is variability in each custom use case, our ML Solutions Lab can help our customers with scoping POCs and production deployment.

## Q6 | What are some examples you're especially proud of?

I'm particularly proud of the really cool work that we've done with professional sports leagues such as Major League Baseball (MLB). For example, we ideated with key MLB stakeholders on a compelling capability for the television audience. Once we decided upon a model that identifies a runner's chance of success for stealing second base, it took us three weeks, starting on September 1, 2018 to prepare the data, train, evaluate and deploy the model.

We leveraged Amazon SageMaker as a key part of this effort, including deployment into production in the cloud. SageMaker provided the requisite sub-second response time for making predictions (called inferences), and, using an auto-scaling cluster, it provided both the high availability and high performance that MLB needed. You'll be able to see this in action at future MLB games, including the 2018 World Series.

## ML IN PRACTICE

## Q7 | What advice do you have for companies looking to embrace ML?

We recommend focusing on how you can deliver a better experience for your customers and identifying the business and operational outcomes desired — and then leveraging ML to meet those needs.

To help realize this potential, culturally, there needs to be an acceptance that ML is an important part of business and operations. Frequently, enterprises have different IT groups, each of which has its own area of expertise or business units that they support. Some ML initiatives may require information from across these domains, so it's important to understand the data strategy and where ML will fit in.

## Q8 | What role does cloud computing play in ML?

The cloud has really done some amazing things for machine learning, because you can try out capabilities. You can do proofs of concepts. If you run into a problem, no worries. You just tried it out. There's no hardware, and you can scale up or down. You have the data there, so you can try other approaches if one doesn't work. Because you can quickly scale up significant amounts of capacity, using ML-tailored Accelerated Computing instances, you can train vast amounts of data in minutes to hours, and that would've taken days in the past.

**AWS TECH**

### Amazon SageMaker: A Powerful ML Platform

When you're ready to test and deploy your own AI solutions, AWS offers an unmatched set of infrastructure and machine learning services. In fact, more machine learning is built on AWS than anywhere else. And Amazon SageMaker is a great place to start. It's a fully-managed platform that enables developers and data scientists to quickly and easily build, train, and deploy machine learning models at any scale. Amazon SageMaker removes all the barriers that typically slow down developers who want to use machine learning by providing a variety of prebuilt algorithms for common business functions, and it supports virtually all popular ML frameworks, such as Apache MxNet, TensorFlow, PyTorch, Cafe2, and Chainer.

"We're trying to put machine learning into the hands of every organization and build out services so that we can help with the undifferentiated heavy lifting of machine learning."

## Q9 What makes ML from AWS unique?

We're trying to put machine learning into the hands of every organization and build out services so that we can help with the undifferentiated heavy lifting of machine learning.

If you look at our ML services stack, you can see that there are many different levels where you can employ AWS services. At the very highest level, we have API-based machine learning services such as Rekognition, Lex, Polly, Comprehend, Translate, and Transcribe. These services are very easy to use and do not require a data scientist.

For customers that want to go a layer deeper, we have infrastructure services, such as SageMaker, that make it easier to build, train, and deploy capabilities into the cloud. SageMaker is an amazing infrastructure service that makes it exceptionally easy to run widely used framework software such as Apache MxNet, TensorFlow, PyTorch, Cafe2, and Chainer. Also, SageMaker includes built-in algorithms that have been optimized and can improve performance by 10x over what you can run elsewhere.

We work closely with companies like Intel to ensure that the hardware powering that infrastructure is custom-built and optimized for the best performance with the most popular frameworks. Also, SageMaker makes it easier for our customers to deploy ML capabilities into production. With click of the button deployment for scalable, multi-availability zone environments, we've taken the IT complexity out of deploying to production.

## Q10 If you were in the shoes of business leader just starting out, what would your next steps be?

First, I'd look for ways to be data-driven in helping customers, optimizing operations to be most efficient, and maximizing financial outcomes. Second, I'd make sure that we're building the people capabilities within our organization to leverage ML, including hiring or training data scientists, where needed. Third, I'd look to identify or establish our organization's approach to data, including governance and repositories, and scoping the IT to support ML in the enterprise. There's no way for businesses to ignore a game-changing technology like ML, and I'm very proud to be part of a team that helps answer questions and move these initiatives forward.

To learn more about ML from AWS and Intel, please visit **ml.aws**

### Intel Innovation in ML

Intel and Amazon have been collaborating closely to drive innovation in AI and ML, bringing you access to high-performing infrastructure, plus machine learning model training tools and inference engines optimized for speed and efficiency.

For example, AWS C5 instances, powered by custom Intel Xeon Scalable processors, are purpose-built to handle the most complex ML workloads. Large-scale deployments can run using platform services like Amazon SageMaker, which is optimized for performance on Intel-based EC2 instances. Deep learning engines from AWS with the latest Intel technology are increasing machine learning performance by over 100x.