O'REILLY®

Compliments of ORACLE

Getting Started with Machine Learning in the Cloud

Using Cloud-Based Platforms to Discover New Business Insights

Alice LaPlante

REPORT

Getting Started with Machine Learning in the Cloud

Using Cloud-Based Platforms to Discover New Business Insights

Alice LaPlante

Beijing • Boston • Farnham • Sebastopol • Tokyo

Getting Started with Machine Learning in the Cloud

by Alice LaPlante

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Jonathan Hassell
Development Editor: Melissa Potter
Production Editor: Deborah Baker
Copyeditor: Octal Publishing, LLC
Proofreader: Arthur Johnson
Interior Designer: David Futato
Cover Designer: Karen Montgomery
Illustrator: Rebecca Demarest

September 2019: First Edition

Revision History for the First Edition

2019-09-16: First Release

This work is part of a collaboration between O'Reilly and Oracle. See our statement of editorial independence.

978-1-492-07275-1

[LSI]

Table of Contents

Getting Started with Machine Learning in the Cloud

Executive Overview

"Machine learning" is a term that we hear virtually everywhere today. Everyone, it seems, is getting into it, and for good reason. Companies are gaining competitive advantage, delivering better customer experiences, increasing revenue, reacting more swiftly to market shifts—all by applying machine-learning techniques and technologies to the data they already possess.

But challenges abound as well. Data scientists with the necessary know-how are scarce, and they sometimes work in organizational silos rather than in collaboration with other stakeholders in the technical and business groups. It can be difficult to integrate machine-learning models with existing or new applications or processes. And a host of security issues in this area have yet to be resolved.

In this report, we go over the types of real-world applications for machine learning that are delivering successes today. We review the benefits and challenges alike of deploying machine learning. We also explain why most organizations are currently implementing their machine-learning projects in the cloud. And we provide you with a list of best practices that early adopters have found useful in their machine-learning deployments.

Introduction: The Data Opportunity

Think of all the data that is collectively accumulating across different areas of the enterprise. Terabytes and terabytes of it. Just imagine if you could really put this data to work for you.

Imagine predicting with an extraordinary degree of accuracy how much of Product A you'll sell next quarter. Knowing exactly which customers will deliver the highest lifetime value. Identifying and fixing inefficiencies in your back-office operations.

This is all possible, if you are taking advantage of your data effectively, and if you are—to use the latest vocabulary—data driven. But recent research by NewVantage Partners finds that most large companies are experiencing problems getting there. A full 69% of IT professionals surveyed said they have failed to create data-driven organizations. Worse, we're apparently heading in the wrong direction. The percentage of organizations calling themselves "data driven" has actually declined annually in recent years—from 37.1% in 2017 to 32.4% in 2018, and down to 31.0% in the latest survey (as shown in Figure 1).

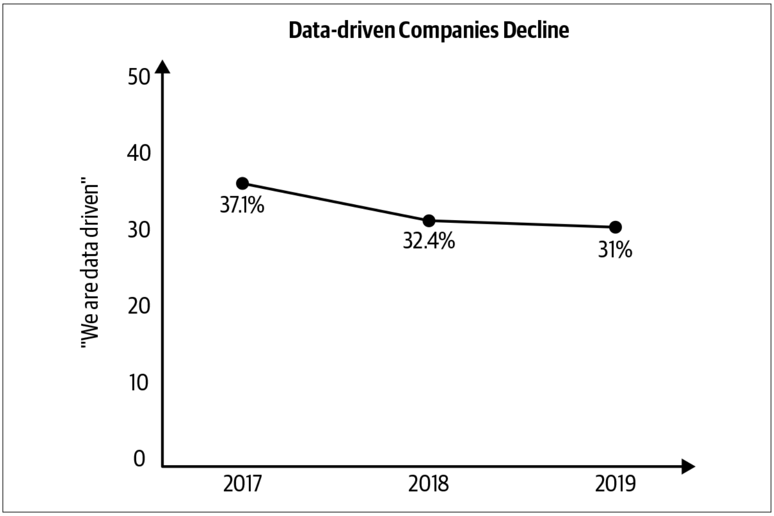

Figure 1. Data-driven companies in decline

Machine learning promises to reverse this negative trend. An increasingly popular technology that falls under the broad umbrella of artificial intelligence (AI), machine learning is predicted by

McKinsey Global Institute to create an additional $2.6 trillion in value by 2020 in marketing and sales, and up to $2 trillion in value for manufacturing and supply-chain planning activities. If you haven't started experimenting with machine learning to tame your data, you could quickly fall behind the curve. In the next section, we give some real-world examples of how machine learning is already yielding often hefty returns on investment (ROI) to companies.

Machine Learning Solves Real-World Problems

Machine learning is paying off for early adopters. A full 82% of enterprises adopting machine learning and AI have gained significant financial advantage from their investments, with an impressive median ROI of 17%, according to Deloitte. Here are some examples of machine learning as applied to challenges businesses face every day.

Recommendation engines

Anyone who has bought something on Amazon or streamed a video on Netflix is familiar with recommendation engines. These are really exercises in personalization. Machine learning helps companies detect patterns based on what customers have purchased in the past, what they've clicked, or even where they've hovered on a particular screen. They use this information in combination with other data—for instance, the customer's age, gender, or geographic locale—to recommend other products or services that they might like, personalized to their tastes—as perceived by the machine-learning algorithms powering the recommendation engine, of course.

Recommendation engines are often expanded with machine learning to predict click-throughs, so businesses know who is most likely to click on an ad. They also learn when to offer timely discounts, including knowing the precise times to send promotional codes to customers so that they will make purchases.

Fraud detection

Machine learning is also being used by credit card companies and other financial-services organizations to detect fraudulent uses of credit cards, identity theft, and account takeovers. And it's not just financial companies that need to be wary of fraud these days. All transactions by online retailers are a potential opportunity for fraud.

One example of a company that uses machine learning to help with fraud detection is Danske Bank. Based in Denmark, but serving the entire Nordic region of Denmark, Finland, Norway, and Sweden, this 145-year-old bank has made fighting fraud a top priority.

Danske Bank's machine-learning fraud initiative was made urgent by changes in customer behavior. Traditionally, most Danske Bank customers had transacted business in person, in branch offices. But today, virtually all interactions are digital—completed through call centers, ATMs, phones, computers, or other mobile devices. Prior to deploying machine learning, the bank had a poor fraud-detection rate—just 40%—and its legacy fraud-detection system was generating 1,200 alerts a day, 99.5% of them false positives. After installing a new machine learning–driven system, the bank cut false positives by 60% and increased genuine alerts by 50%, and could focus its resources on the cases that truly required the attention of its security staff. Danske Bank did this by integrating deep learning software with graphical processing unit (GPU) appliances optimized for AI-based deep learning. Data scientists wrote an analytic model that identified potential cases of fraud while intelligently avoiding false positives. Additionally, most operational decisions were shifted from human workers to the AI-based system.

Retention modeling

Retention modeling systems predict whether—or when—a customer is likely to abandon a brand. This information is extraordinarily valuable to businesses, giving them the opportunity to take appropriate actions to retain at-risk customers and, in the process, increase their lifetime value. Knowing who your most valuable customers are is particularly critical in the online retail space.

Donatos Pizza, a family-owned pizza parlor based in Ohio that has grown from a single location in 1963 to more than 160 franchises today, noticed that its clientele was dropping off. Using machine learning, Donatos was able to identify customers it was at risk of losing and consequently take immediate action to retain them, such as offering coupons for free food items. Within months of applying machine-learning algorithms to its database of customers' past behaviors, Donatos was able to retain 45% of these at-risk customers, and it achieved a 15% improvement in overall customer retention. So successful was the program that Donatos made it obligatory across all of its franchised locations.

Predictive maintenance

Predictive maintenance driven by machine learning is a huge—and growing—trend, especially in manufacturing, aerospace, oil and gas, and other industries that must manage expensive, mission-critical heavy equipment. Rather than a company depending on rough estimates of when a part should be replaced or other maintenance performed on machinery, machine-learning algorithms can predict when a piece of equipment is likely to fail based on a broad range of signals detected and learned from that particular piece of equipment itself.

This ability is extraordinarily valuable. It prevents unplanned downtime due to failures in key equipment, reduces the need to keep excess spare-part inventory on hand, and minimizes the overtime pay of personnel required to fix equipment in emergency situations.

AGR is a leading global oil-well design-and-drilling project-management and field-management company whose services cover the entire oil-field life cycle. It manages hundreds of oil and gas production plants, many of them in remote locations. If broken equipment delays or stops oil or gas production, the costs can range from tens of thousands to hundreds of thousands of dollars in lost revenue. By deploying machine learning and advanced analytics, today AGR can predict when equipment needs maintenance prior to an unplanned disruption, and can take timely action.

Despite early successes at AGR and elsewhere, the use of predictive maintenance and machine learning is still in its infancy. A recent study showed that a plurality of industrial organizations (44%) still rely heavily on simple ad hoc analyses done in Excel, for example; that's significantly more than the percentage of analyses done via machine learning (12%). Much of this difference can be attributed to the complexity of developing predictive maintenance systems for specific pieces of equipment.

Medical diagnoses

In the field of medicine, machine learning is being applied to medical diagnostics. Its capacity to analyze numerous variables in huge volumes of data makes it possible to predict which patients are most likely to contract certain kinds of diseases, and to take preventive measures to slow down or even ward off symptoms.

For example, more than 12,000 children are diagnosed annually with diabetic ketoacidosis (DKA), a life-threatening complication of Type 1 diabetes. Texas Children's Hospital used machine learning to more precisely predict DKA risk in time to take preventive measures. Deploying a machine-learning system resulted in an impressive 30.9% decrease in DKA admissions per year and the formation of an interdisciplinary team to collaborate on high-risk cases as predicted by the machine-learning system.

All of these successes are building a convincing case for machine learning. In the next section, we talk at a high level about some of the most compelling benefits.

Benefits of Deploying Machine Learning

As the real-world examples in the previous section reveal, you can gain many business benefits from machine learning. We discuss these benefits in more detail throughout this section.

Help You Make Smarter, Faster Decisions

Machine-learning models deliver recommendations and predictions faster than you could previously hope for. Given the volumes of data that your connected organization collects every day—data it generates itself, data from operations, data from transactions, and data from partners—it would be impossible to process all of this information and act on it without relying on an AI technology like machine learning. Real-time decision making based on hard data is now possible—giving companies that have deployed machine learning an edge over competitors, many of whom are still, for example, getting their sales data two weeks in arrears.

Help Remove Bias from Decision Making

A machine-learning algorithm has no preconceived notions of how a particular business problem can or should be solved. It doesn't rely on "gut feelings" as do so many business executives, feelings that can lead an enterprise down a wrong path. Machine learning looks at the data and nothing else. It allows analysts to uncover structure in the data and derive insights so that you can make better business decisions.

Of course, because any "learning" that the machine-learning system does will be based on the data, it is essential that the data is not biased. Human beings have their blind spots, and the data they generate and collect and analyze can reflect that. In several notorious instances, machine-learning systems that had been used to make important business decisions (such as hiring for strategic jobs) were later revealed to be incredibly biased because they were fed biased or incomplete data.

With the right data, machine learning is also able to adapt to new knowledge or trends more swiftly than people can. So, in an environment where markets or customer behaviors are shifting rapidly, machine learning is a very appropriate tool.

Leverage All Your Data

Data is the name of the game in machine learning. Unlike traditional programming models, machine-learning models can be trained on virtually unlimited amounts of data to predict valuable outcomes and drive business decisions. For example, you might want to use customers' online behavior data to drive your advertising budget—to tell you where to invest your digital ad dollars. In such a case, the model will be hungry for any data you can feed it—structured, unstructured, and everything in between. Then the model uses this new data to continuously adapt for advertising strategy. By taking data from your website—who clicked on what—and combining it with data on who clicked through on your Facebook ads and with what you know about your existing customers, you can predict who is most likely to purchase based on ads and target your advertising budget accordingly.

Deliver a Better Customer Experience

Because machine learning helps you understand your customers much better than you could previously—goodbye to the crude market segmentation techniques of yesterday—you can personalize experiences to meet their particular needs and preferences. According to an IDC survey, AI-based customer relationship management could boost global business revenues by $1.1 trillion from the beginning of 2017 to the end of 2021. This boost will be driven mostly by increased productivity ($121 billion) and lowered expenses due to automation ($265 billion).

Streamline Operations and Reduce Costs

Until fairly recently, our ability to automate business processes was limited to those processes that were predictable and repetitive in nature—for example, feeding structured data from your accounts payable system into your accounting system. If any part of the process required taking new information into consideration or making a judgment call based on more than simplistic rules, the organization needed to perform it manually. With machine learning, you can now automate processes that are variable and unpredictable because the system learns as it processes data in real time. For example, you can use machine learning to look at unstructured data coming from a wide variety of invoices and automatically assess and pay the ones that match your records, while routing problematic invoices to a human to resolve. This is revolutionizing both front- and back-office activities, speeding up operations, improving employee productivity, and saving money.

Now that we've covered the benefits, it's time to dig a little deeper into the details. A major question that comes up with AI and machine learning is, where should it be deployed—on premises or in the cloud? We explore this question in the next section.

To Cloud or Not to Cloud?

One of the first—and most important—decisions you will need to make when considering machine learning is, where do you locate your machine-learning models: on premises or in the cloud? Increasingly, companies are choosing the cloud. A recent survey by Deloitte Global found that 2019 is the year when using the cloud for AI—which encompasses machine learning—will dramatically accelerate.

Cost is likely a factor here. The infrastructure required to support an effective machine-learning model was once prohibitively expensive, putting it out of reach for many organizations. The public cloud changed that. Suddenly, more businesses are finding it possible to have complete machine learning–ready environments available on demand, and pay only for what they use.

According to this same Deloitte study, among companies that adopt AI technology, 70% will subscribe to cloud-based enterprise software, and 65% will create AI applications using cloud-based development services. By 2020, an estimated 87% of organizations that use AI software will use cloud-based AI software.

The top cloud computing platforms are all betting big on democratizing AI. All have made significant investments in machine learning and are announcing new services every few months. And many are beginning to reap the benefits of the cloud, which we discuss in the next section.

Benefits of the Cloud

The first benefit of the cloud is *elasticity*. The compute and storage requirements for machine learning are extensive because of the sheer volume of the data that needs to be stored and processed. Plus, machine learning can be enormously variable in its demands, requiring a lot of resources during one hour and far fewer the next. The cloud is the perfect place for large and variable workloads, because you can scale up—and down—as required. It's also more cost effective than on-premises because you only rent the resources you need when you need them. No need to overprovision on-premises infrastructure to meet your maximum projected requirements for peak workloads.

A second benefit is that you can *take full advantage of data* you previously couldn't—from data warehouses, software-as-a-service (SaaS) applications, and on-premises databases. In the past, it was difficult to create a dedicated, on-premises environment (such as a data lake) to hold all this data, due to complexity, cost, and data variability. Today, you can put your data lake into the cloud, and it will hold anything and everything you need until you are ready to analyze it or run it through machine-learning models.

Agility is the third major benefit of using the cloud for machine learning. Because you can provision compute and storage instantly, you can move on new ideas more swiftly and adjust to market shifts more seamlessly. You also get faster access to the latest technologies because you don't accumulate the technical debt that you would if you were buying, not renting, all the required resources.

For example, you can take better advantage of the latest advances in GPU technology in the cloud than if you'd invested in a previous generation of the technology. And because whatever is stored in the cloud is accessible to anyone anywhere, the cloud can truly democratize data-driven decision making, leading to much more agile business operations. As an added bonus, the major cloud providers offer machine-learning tools and capabilities that don't require a data scientist; business users can easily learn these tools, which streamlines machine learning and makes it easier for everyone. Tools like Oracle Machine Learning Notebook, Amazon SageMaker, Google Cloud ML Engine, and Azure Machine Learning services offer options such as prebuilt algorithms or trained models for businesses to easily use.

Finally, you get *operational efficiencies* from the cloud because you don't need to worry about maintaining hardware and patching software. All that is done for you by the cloud vendor. This frequently leads to cost savings over operating an extensive on-premises machine-learning operation.

The next section looks at the machine-learning life cycle: the step-by-step process by which you create, deploy, and retire machine-learning models.

The Machine-Learning Life Cycle

The machine-learning life cycle is a process that defines each phase for deploying machine learning to get practical business value out of your data. There are six sequential phases, each of which is critical for deployment success. It's important to understand that the life cycle is iterative: when you get to the final phase, it's time to begin thinking about starting all over again.

Phase 1: Establish a business goal

You need a tangible goal. Whether that's to increase revenues, enhance the customer experience, streamline operations, or otherwise create value for your organization depends on your reasons for turning to machine learning in the first place. Take online retail. You might have the goal of identifying which customers are likely to leave so you can take preventive action and improve the lifetime value you derive from them. But too many organizations fail to start with a business goal. Instead they say, "We have all this data. What can we learn from it?" If you approach it that way, you will end up spinning your wheels and wasting time and money. It's always good to start with a smaller and more easily achievable goal (see "Step 1: Formulate a Concrete, Achievable Goal" on page 12).

Phase 2: Acquire, clean, and prepare data

Now you need to gather and prepare all relevant data. In the online retail case outlined earlier, this would mean consulting with experts within the business as to which attributes might be most important to predicting whether a customer might stop doing business with you. This can range from data on gender and physical location to historical data of price fluctuations, number of dropped calls, changes over time, or myriad other variables. These are called *features*, and the process of choosing them is called *feature engineering*, or using in-depth knowledge of the data to create features that will make machine-learning algorithms work. Then you need to scrub and clean the data—nothing is more important in a machine-learning project than reliable data—and get it into a format suitable for analysis.

Phase 3: Build and train models

There are two types of machine-learning models: *supervised learning* and *unsupervised learning*. With supervised learning, you start by deciding what your "target variable" is. (This is not the case with unsupervised learning.) In our retail example, it would be whether your customer continues buying from you or not. Data scientists and *citizen data scientists* then run machine-learning algorithms on the data. These algorithms build models—sometimes hundreds of them—based on different algorithms or machine-learning approaches. These models are then trained by the historical customer data that has been collected and prepared. Finally, you take the trained models and run them on new data to see whether they correctly predict customer churn. You pick the model or models that return the best results.

Phase 4: Evaluate and understand your model

One of the most difficult tasks of machine-learning projects is to evaluate the performance of a model and understand its inner workings. Even for data scientists, it can be difficult to understand why a model is making the predictions that it is making. But before you deploy a machine-learning model, you need to understand which of the many data features you've collected affect the outcome. This is an important activity for any project and can even be mandatory in heavily regulated industries such as banking and health care.

Phase 5: Deploy and operationalize the model

The next step is to deploy the model into production. This is often the most difficult step because of disconnects between the technologies used to build the model and the production technologies. In fact, by one estimate, 87% of models never make it into production. For successful completion, this step requires intense collaboration among infrastructure professionals, data scientists, data engineers, domain experts, end users, and management.

Phase 6: Monitor and retire the model

This is the "last mile" in the model life cycle and is often neglected, despite being a key aspect. You must continuously monitor the model—in real time, if possible—to determine how well it is performing. The accuracy of models degrades over time. They become less effective. They need to be refreshed or replaced. Unfortunately, many early users of machine-learning models do not do this, and as time goes by their machine-learning deployments return less and less value. Or they do it at fixed intervals, when ideally it should be done continuously, in real time.

The reason for this is that the world is constantly changing, and thus the data you are collecting reflects this. Customers change their preferences. Fraudsters actively monitor what works and change their attack modes to avoid being identified. As a result, the new data coming in could be very different from the data that was used to train the model. Thus, the goal of monitoring is to identify when the model becomes subpar in performance. Then it's time to return to Phase 2.

How to Get Started

After you understand the machine-learning life cycle, it's time to get started on a machine-learning initiative. Much like the process of actually doing machine learning, there are steps to boosting the success of your machine-learning deployment. Here are five basic steps to getting started.

Step 1: Formulate a Concrete, Achievable Goal

You need to have a vision that is realistic—a solid business goal—and then you need to "think small" within that vision to get going. The trick is to be extraordinarily specific about the business problem. For example, rather than "I want to know whether a customer is going to churn," you should say, "I would like to know within three months of renewal whether a customer is going to cancel their subscription." Or, if you are a hospital wanting to prevent heart-attack fatalities, you might say, "I want to know the vitals and symptoms of the patient three months before she had a heart attack."

Step 2: Assemble the Team

The next step is to understand that putting together a machine-learning team is not just finding data scientists—as difficult as that will be in and of itself. You need to be able to put together a multidisciplinary team. That means not just machine-learning engineers and data scientists but also data engineers, application developers with DevOps engineers to support them, and business analysts to take the insights generated by the models and apply them to real-world challenges. By themselves, data scientists generally don't bring models to production. They surround themselves with other professionals to facilitate that.

Step 3: Achieve Quick Wins That Show the Value of Machine Learning

The most important thing is to think small so that you can prove immediate, tangible value. Don't try to impress the boss with how you're going to take all unstructured data from Twitter and combine it with sensor data from the internet of things (IoT) to do something risky and time consuming that ultimately fails. Instead, think of some finite problem that nevertheless is holding back your organization in some way. Perhaps your hospital is readmitting too many patients. Or not enough people are moving from the freemium to the premium version of your product. Identify what you'd like to accomplish, set an attainable goal, and reach it. You can then achieve quick wins through discovering new insights and potential cost savings, and that's very valuable. But keep in mind that machine learning isn't truly considered valuable in the eyes of the business until it starts to improve revenue, which is related to the coming step of operationalization.

Step 4: Evangelize to Get Executive Support

When you have proof that machine learning is working for you, loop in senior management. Executives focused on digital transformation and transitioning the company to becoming data driven will be interested, as machine learning is a proven way of achieving both those outcomes. Moreover, an executive sponsor is likely needed to okay budgets for expanded machine-learning activities.

Step 5: Operationalize/Productionize Machine-Learning Initiatives

Finally, you need to ensure that your machine-learning initiatives are consumable by other parts of the organization. If your model is not consumable, it does not exist. This means getting the model into production. However, some of the steps between developing a machine-learning model and producing it require engineering skills. But many data scientists don't possess the skill set to complete a complex engineering task like deploying a REST API endpoint. And the last thing you want is for the machine-learning models to stay on the laptops of the data scientists.

Some of the new cloud-based data-management and machine-learning platforms make it easy for data scientists to build and operationalize models by themselves, so they can directly put them into production or undergird larger enterprise applications without involving too many other disciplines. However, because data scientists typically prefer languages like Python or R, whereas production requires a model to be in Java, all too often you're looking at lengthy time frames for getting models into production.

Challenges of Getting Started with Machine Learning

Poor data quality and availability are some of the biggest machine-learning challenges. No matter how well built the model, its output will only be as accurate as the data that you feed it. Deloitte recently surveyed 1,900 executives who have begun to use AI technologies, including machine learning, and found that 38% of them ranked data access and data quality issues as a top-three challenge. Moving data around from databases to tools costs time, money, and resources, and it's best to minimize data movement if possible.

Ensuring collaboration between data scientists and other relevant parties can also be a challenge. Data scientists can sometimes operate in silos. They work on their laptops and don't share their code, so if a data scientist leaves the company, it's a nightmare to figure out where a model came from or where the source code is stored. Without a framework for oversight and governance to ensure that code is uploaded to a central repository—or without strict versioning and labeling of models—your company could get into trouble down the road.

Then there's the challenge of sometimes needing to rewrite models to integrate them into applications or reports, or to share insights and predictions with analytics products such as Tableau, Oracle Analytics Cloud, and other solutions. This is getting easier thanks to the new cloud-based hybrid data management and machine-learning platforms that are now available, but the operationalizing phase is still the most significant roadblock for many companies.

Finally, security is always a challenge, especially when you're exporting data into external solutions or analytics platforms. How do you know that somebody is not trying to hack into that data or change values in it? Security is a whole dimension that most people have completely ignored when it comes to machine learning.

So how to meet these challenges and reap the promised benefits of machine learning? We've collected a portfolio of best practices from industry experts, which we lay out in the next section.

Best Practices for Machine Learning

Early adopters of machine learning have picked up some experience from their initial forays into the technology. Here's their advice.

Promote Intense Collaboration

No data scientist should be an island. Organizations that pull together interdisciplinary teams of business managers, analysts, data engineers, data scientists, and DevOps personnel, among other employees, will succeed the most with their machine-learning initiatives. These professionals need to collaborate on ideas, share insights, and give and accept constructive criticism in the proper spirit. Each of them possesses specialized knowledge that others don't have, and if they don't interact with one another, projects will fail. For example, the business analyst—often the most overlooked member of the team—has valuable domain knowledge about, for instance, finance, sales, or HR, that provides essential context to the data scientists building the machine-learning models.

Get Started Now—but Don't Try to Boil the Ocean

Start something. But start small. Don't try to solve the problem your company has never been able to solve before. You might want to learn simply how to sell more product. Or identify who isn't going to renew their subscription. Or nurture the employees most likely to be future leaders.

Begin with a simple model with few features and a smaller set of dependencies. Deploy it, and then as you iterate, slowly add more features. The more features you have, the more downstream dependencies you will get. You might want to connect to different data sources, or you might want to add new attributes and engineered features from your databases. A column in your database can be the outcome of another completely separate data transformation pipeline that could fail. If you have a lot of features, and one of the tables is not filled out properly or one of the fields is not correctly set up, your model is not going to perform as expected in production. So reducing the number of dependencies you have on data is generally better for your first-time models.

Additionally, avoid using this initial model to predict outcomes on all of your new data. Process incrementally after you have convinced yourself that the model is behaving correctly in production. You certainly don't want a bad model to affect all your customers. Start with a small subset and progressively increase the size of that subset. For example, if you have a machine-learning model that's going to change the user experience (UX) on your website, you should feed it only a small portion of traffic. Then, if the model is doing well, you can increase traffic incrementally.

Identify the Correct Success Metrics

We've already mentioned that you need to start with a concrete business problem. Beyond that, you need to make sure that the success metrics you establish are the best ones—and that they are precisely defined. If you want to increase overall revenues, do you want to measure that across all your product lines? Or only for those that have some price elasticity in them? Or say you want to identify your "best" customers. How do you define "best"? Most revenues sold to them overall? Lowest number of support calls? The more specific you are, the more likely your model will deliver value. Table 1 presents a few more examples that look at this more closely.

Table 1. How to make your metrics more specific

Poorly defined metric	More precise metric
Increase utilization of resources	Reduce empty seats per flight by 5% and save $1 million per quarter
Identify my best customers	Identify customers with the greatest potential to become a top 10% customer for estimated lifetime value
Fight fraudulent credit card purchases	Minimize false positives for declined transactions
Predict who is likely to buy Product A	Predict who is likely to click an ad for Product A

Democratize Access to the Data

Empower the people who need to use the machine-learning outputs with direct access to the models and data. Data scientists and data engineers have traditionally been loath to do this. Accustomed to being the gatekeepers of data and responsible for keeping it clean and error-free, they have been fiercely protective of such resources. They didn't want "amateurs" messing with them. With today's smarter and better hybrid data management and machine-learning platforms, there are no longer any excuses. Superior platforms will allow you to do all your predictive modeling and machine learning inside the data management platform. This accelerates making predictions and generating new insights, especially if you're able to move algorithms to the data instead of moving data from the database and dealing with the resulting complexities. Rather than working separately, and almost at odds, data analysts, data scientists, data engineers, and IT can work together in a controlled fashion, and work better and faster.

Automate as Much as Possible

There are only so many hours in the day. So automate whatever can be automated. Create dashboards that are automatically populated and refreshed so that you don't need to constantly run custom reports. Incorporate machine-learning insights directly into enterprise applications.

Then there's automated machine learning (AutoML). AutoML is an emerging technique of automating aspects of applying machine learning to business challenges. In the traditional machine-learning life cycle (see "The Machine-Learning Life Cycle" on page 10), data scientists must spend a good amount of time preparing, preprocessing, and feature-engineering the data. They then need to select the appropriate algorithm(s) to build and optimize the best machine-learning model for the job. All of these tasks require highly specialized expertise. AutoML can automatically perform feature selection and create machine-learning models that are often superior to those crafted manually by data scientists. In other words, you simply feed data into a "black box," and it returns the best possible machine-learning model to meet your goals.

Although in its infancy, AutoML is promising to significantly speed up the creation and deployment of machine-learning models in today's businesses.

Don't Abandon Your Common Sense

If the results of a machine-learning model don't pass your sniff test, be sure to follow up. Check and recheck the data for common sense anomalies, and make sure you understand why you're getting the results you are getting.

Support Multiple Models at the Same Time

Make sure that you can support multiple models in production at the same time. This means trying different algorithms to achieve your aims and running them in competition to see which ones return the best results. You can experiment with them, run A/B tests, or take a "multiarmed bandit" approach to model deployment. It's also important to note that although you might think you found the best model during training, that model might not perform as well in production. Divide traffic instead among multiple models, putting for example 10% of traffic through each one of your 10 models, and then changing that allocation depending on how the models are performing in production. In short: never put all of your eggs in the same basket.

Do Careful Version Control of Your Code for Reproducibility

For reasons of reproducibility, always ensure that your code is version controlled at all times. That's a requirement. The data should also be versioned. This way, you'll always know exactly what steps were involved in building your dataset, and you'll know exactly what data went into training your model.

You should also track all the dependencies—the libraries, the environment, and the language that was used to train your model. These might sound trivial, but in practice, it's really important that you track all of these things so that if you need to reproduce your results, you can do so easily. This is not discussed often, but it's an essential aspect of a successful machine-learning initiative.

Reproducibility is in fact required in a lot of regulated industries. You must be able to reproduce a model from scratch. You need to know exactly what data was used and what steps were involved in the transformation of the data. You need to know what algorithm was used. And you need to be able to audit the entire training process—in fact, the entire machine-learning life cycle, from data acquisition all the way through to model deployment, needs to be reproducible.

You need to monitor model performance, rebuild models on the more recent data, see new attributes and new "engineered features" from your data that can possibly improve model accuracy, and repeat this machine-learning cycle for incremental improvements.

Monitor in Real Time

And then, after you've deployed your model to production, you need to continuously monitor it. Not all teams do this, much less in real time, but it's critical. You want to see the memory consumption of your production environment. You want to know the CPU consumption; you want to know how many errors your model has returned. That's all about the general health of your production environment. But even more important is the other element to monitor: *concept drift*, which is about being able to stage the model predictions and see whether the model is drifting or whether its performance is slowing down over time. A broad set of diagnostics and tools are available to help data scientists discern whether a model is underperforming and needs to be retrained or even to be redesigned from scratch.

Want to see machine learning in action? In the following section, we demonstrate how two companies successfully deployed machine learning to address pressing business challenges.

Case Studies

Machine-learning successes are beginning to accumulate out in the real world. This generally occurs when organizations realize they can't get full value from their data through traditional databases or analytics solutions. Here are two case studies of companies that have achieved significant gains with their machine-learning initiatives.

CaixaBank Case Study

CaixaBank is the current leader of the Spanish retail banking market. It has almost 16 million customers—with 6 million of them online. It also has the most extensive branch network in Spain, with more than 5,000 offices and 9,000 ATMs, and assets worth 342 billion Euros. CaixaBank is organized as an integrated financial group, with a banking business, insurance activities, and investments in international banks. It is growing both domestically and internationally, thanks to its proven experience in banking sector investments and its fiscal savviness.

In recognition of its digital transformation strategy, CaixaBank has become one of the highest-rated banks in the world based on the quality of its digital products and services. In 2019, the company was recognized as the "Most Innovative Bank in Western Europe" by the US magazine *Global Finance*. It has also been acknowledged as the "Best Private Bank for Digital Client Communication" at the Wealth Tech Awards held by *PWM* magazine of the Financial Times Group.

Data mining is not a new activity for banks in general, and CaixaBank in particular. CaixaBank began using data to increase sales and offer better customer experiences more than two decades ago. But approximately 10 years ago, the entry of new digital players in the financial payments market gave CaixaBank a wake-up call. CaixaBank realized it had to be prepared to use the same technologies as these digital competitors, and they turned to Oracle's suite of machine-learning products to help them outpace the competition.

Democratizing machine learning

Banks like CaixaBank have been using analytics models for many years, but machine learning is helping to improve the accuracy of those models.

Previously, CaixaBank's models were built using statistical techniques and derivations. But with machine learning, it can achieve more efficiency in many parts of the business, not only in the expected ways—like risk or commercial lending—but also by optimizing product prices, performing better liquidity forecasts, improving employee productivity, and preventing money-laundering activities.

"Machine learning can be used in many, many ways throughout the bank," says Xavier Gonzalez Farran, the director of big data analytical tools at CaixaBank, in Barcelona. "In fact, we are already using it in almost all departments and all of our business lines."

Gonzalez says the bank is passionate about spreading what he calls "machine-learning literacy." This means helping employees understand what machine learning is capable of, and applying it to the real-world challenges they face in their jobs. "Today, when a problem appears in some part of the business, our goal is for employees to immediately think, 'Okay, I can apply machine learning to solve this problem.'"

Significant profitability gains

One example of how CaixaBank used machine learning was in performing risk analyses when granting loans. Before this project the bank used statistical derivation models. But using more sophisticated machine-learning algorithms, specifically gradient boosting machines (GBMs), it managed to improve the accuracy of its model by 7%. This led to an increase in profits on its loan portfolio—a huge win for the bank.

CaixaBank is also using a machine-learning algorithm to manage the issues that arise daily with direct debit transactions. In Spain, it is common for consumers and businesses alike to pay their bills for electricity, water, or other essential services through their bank accounts. But in a typical bank, thousands of issues will arise daily related to direct debits, from wrong account numbers to someone not having enough funds in their account to pay a bill. Before machine learning was deployed, all these issues had to be managed manually by branch employees.

So the bank started to train the machine-learning algorithm with the thousands of historical decisions that had already been made by human workers when solving direct deposit snafus. Ultimately, the algorithm was able to reproduce the decisions of human workers with a 99% accuracy. The bank calculated that it is saving many employee hours. Not only is this a huge cost savings, but branch employees can now dedicate themselves to providing better financial advice to customers.

"It's turned out to be more profitable for us to reassign our employees than to force them to spend a mountain of hours on routine, back-office tasks," says Gonzalez.

Waiting to move to the cloud

Regulators in Europe are not yet allowing EU financial institutions to move to the cloud, but Gonzalez thinks that this will change because the European financial services industry is united in telling the regulators that this must change. "Machine-learning algorithms need very powerful servers," he says. "If you're on-premise, it can be difficult to grow quickly if you need more hardware or power for your computation. The cloud is the perfect business solution for that."

Eventually, the bank will be moving to more sophisticated algorithms like deep learning and neural networks, and will need even more computation power, "and I think the cloud will eventually be the solution for all these types of problems and projects," says Gonzalez.

Quick wins needed

CaixaBank needed some quick wins to gain management support, and Gonzalez said the ease of using Oracle's machine-learning tools instead of relying upon DIY tools helped it achieve them.

"These projects are quite expensive, and the investment at the beginning is quite high," says Gonzalez. "So if you can develop some successful models quickly and easily, you convince top management to give you more support."

Four pointers for organizations getting started with machine learning

For organizations just starting on machine learning, Gonzalez has four pieces of advice:

Spread machine-learning literacy through your organization.
: But at the same time impose robust governance in methodologies and continuously monitor these models. This is important because the number of machine-learning models is going to increase very quickly.

  "Within the next year, I think that we will see many companies have hundreds or perhaps thousands of models in production, and if you don't have good governance, you can lose control of all that software and it's very dangerous," says Gonzalez.

Build a collaborative culture.
: This will boost your capacity to generate new models based on the previous work of other users, and will accelerate your machine-learning deployments. Unfortunately, data scientists are not so good at sharing knowledge, and prefer to develop their models alone, so this is something "that you have to make a lot of effort to achieve," says Gonzalez. Enabling collaboration between different departments is just as important, so setting up a good model repository and a place to store group documentation materials is essential.

Ensure the quality of your data.
: When machine learning is truly democratized, sophisticated algorithms are now more accessible and can be used by more people in the bank. But it's important that these algorithms use good data—both when they are being trained, and when they are deployed.

Think about how well-integrated your data and solution are.
: Choosing a machine-learning solution that's completely integrated within the database makes it very easy to deploy algorithms without having to manage different technologies. And running models in the same place where the data resides means not having to move that data to another server. That also makes it possible to embed machine-learning algorithms in SQL processes or integrate execution of the model within a business process.

  "It's critical to focus not only on the algorithms and the machine-learning models, but also on the data model," says

Gonzalez. “In many ways, preparing data is one of the longest and hardest tasks that data scientists have.” When CaixaBank started its big data initiative six years ago, it integrated and stored all its data in one data warehouse. “But all the sources weren’t homogenous, and it wasn’t easy to mix data from one data set with another,” he says. So the bank set out to build a corporate data model that reflected the bank’s business. It is where the bank applies all its data quality strictures and data governance procedures to ensure that data ingestion and data transformation are done to the highest standards. “We invested a lot of resources in building this model, but it has helped us get the good results we have experienced,” he says.

DX Marketing Case Study

DX Marketing (DXM) is an award-winning marketing insights company that utilizes unique consumer knowledge to inform businesses’ strategies and deliver quantifiable ROIs on their marketing spending. Through the diverse set of partnerships that it has built, DXM gives its clients access to the most comprehensive data on individual consumers and their digital behaviors. This lets them see what consumers actually do, what influences them, and what ultimately leads them to buy. DXM’s proprietary methodologies affect the entire marketing life cycle, from segmentation to market potential, consumer engagement, acquisition, and, finally, retention. Established in 2001, DXM has offices in Greenville, South Carolina, and Savannah, Georgia.

DXM’s massive customer data warehouse tracks 830 attributes on more than 280 million records and performs advanced analytics on all this data about consumers based on what its clients want to know about their customers—both current and potential. Using advanced analytics and machine learning, the company’s team of data scientists and marketing experts sift through this massive volume of consumer profiles to quickly and accurately predict likely buyers for a wide range of financial, retail, automotive, education, telecom, travel, and health-care industry clients.

A cloud pioneer

As DXM grew and expanded its offerings to include new capabilities, such as integrating geospatial datasets into its data warehouse, it realized that it didn’t have a suitable infrastructure to scale the amount of data being processed each day. Building its own on-premises infrastructure to house all of its data would require too much of an up-front investment and prevent it from focusing on its core business: marketing. DXM decided to go to the cloud with Oracle, according to Ray Owens, CEO and founder of DXM. Building an on-premises data warehouse would have taken 9 to 12 months, whereas the time to market in the cloud—from spinning up a self-driving Autonomous Database that came with machine-learning algorithms through Oracle Machine Learning, to going into production—took only three weeks. "That was big for us to get to market really quickly, with such robust capabilities," says Owens.

It wasn't just the ability to use a self-driving database that was invaluable, however. It was the fact that DXM could run its machine-learning algorithms in the database. That meant there was no need to export data to a separate cluster (with all the cost, time, and expertise needed to make that work). Moving algorithms to the data instead of moving the data to the machine-learning algorithms greatly improved the company's speed in finding machine-learning insights and then operationalizing them.

The need to interconnect easily with lots of data management platforms

Today, when DXM connects to the data management platforms (DMPs) of its clients, it generates a lot of metadata through consumers' exposures to ads during campaigns—and a typical consumer dataset of millions of campaign records will return a 10-times factor in volume. Clients then feed their data about click-throughs and actual point-of-sale transactions back to DXM's data warehouse, which is able to sort through all the various data points using advanced analytics and machine learning and come up with insights into consumer behavior that otherwise would not be discoverable. DXM estimates that it routinely sorts through 70,000 unique data points—multiplied by millions of customer transactions—on behalf of its clients for each campaign.

DXM also uses machine learning on geospatial data, measuring distance calculations, competitive overlays, and potential trade area cannibalizations for brick-and-mortar locations. It has an instance inside of its self-driving data warehouse in which it stores large volumes of physical mapping data so that it can create maps and do some very robust visualization for clients on their customer and prospect audiences.

Data flows both ways

The data flow is important because DXM stores prodigious amounts of consumer data in its self-driving data warehouse, along with all its clients' customer data. DXM works today with roughly 35 different clients to build predictive models using machine learning and advanced analytics to grow their respective businesses.

Then, of course, while running campaigns, it gets lots of campaign data back: impression logs, exposure logs, and other indications of how the campaign was received by consumers. For example, after building machine-learning predictive models of audiences, it places online ads to show to the selected demographic and then uses the point-of-sale data from its clients to correlate what consumers actually bought with other variables.

"What ends up happening is that business analysts can go into the cloud-based data warehouse and, without the aid of a database administrator, can create scripts and start connecting all this data to the creative ads that are being served," says Owens. The exposure files and responder data continue to come from clients, and all of this goes into the self-driving data warehouse to make marketing efforts more transparent as to which channels are more effective for clients' products or services.

The critical part of the self-driving data warehouse is that DXM can run its machine learning directly in the database itself. There's no need to export data to some separate cluster, with all the cost, time, and expertise required to make that work. In short, DXM is able to bring the machine-learning algorithms to the data, as opposed to moving the data to the algorithms.

What DXM did for a telecom client

For example, for a major telecom firm, DXM took 70,000 touchpoints, drew 15 million prospective customers into its DMP, and attached them to the various data sources that are already in the system. The question the client wanted to ask was: will prospective customers pay for quality technology, specifically cell phones and wireless technology? "We were specifically looking for intent based on attitudinal data about technology," says Owens.

After DXM narrowed down its database of prospective consumers, it was easy to passively observe their relevant behavior in an aggregated and privacy-protected way because customers frequently used their devices to search for a variety of different products and services related to DXM's client's product. "We collected a lot of data, and used machine learning to sort through it and attached the right data nodes to our existing dataset to determine who was most likely to convert for our client," says Owens.

From there, DXM pushed the data into programmatic campaigns, AdWords accounts, social media, and streaming radio. "It's a very, very powerful connection to being able to deliver messaging to a relevant channel for your prospect or customer," says Owens. When the campaigns started to return data, DXM went back to the data warehouse, ran machine-learning analytics through various demographic, channel specific, and geospatial data points in the algorithms that were included within it, and looked at what they'd learned.

A key "aha" moment

For clients, especially those in the digital space, one of their biggest questions is what their return on marketing investment (ROMI) is. For example, if a company is showing a consumer ad every month for six months to a prospect who is showing behavioral intent, how long does it take—and how much does it cost—before that consumer purchases something?

"For this particular client we learned that responses and conversions tended to drop off after four months in the decision cycle," says Owens, who said this was a "huge" insight. To get this, DXM had simply fed all the exposure data into the data warehouse, and one of the machine-learning insights was a correlation that told DXM about the relationship between in-market impressions and the drop-off in purchases.

So rather than spending money on an ad campaign for 6 to 12 months while hoping that consumers would buy their products, this client learned that limiting itself to a four-month window of ads in a marketing campaign where the prospect was showing in-market intent would get optimal results.

Because the clients that DXM work with share their actual transaction data, DXM can use machine learning to calculate how much a business spent to acquire a customer.

For example, the telecom client wanted to reduce the cost of acquiring a customer from $85 to something significantly lower. DXM was able to reduce the cost of customer acquisition to just $36. “We built the predictive model, analyzed the behavior in the DMP, and then throttled the advertising to them when they actually showed actual intent. We then stopped the six-, seven-, eight-, and nine-month campaigns, because they were just wasting money. It turned out that three to four months of ads were sufficient to the right audience,” says Owens.

Another surprising insight

DXM’s telecom client also wanted some way to indicate which prospects were most likely to be able to afford to pay their bills. DXM was already getting a fairly detailed view of the telecom’s current and potential customers, but it wondered whether another variable could contribute to consumers’ scores beyond typical risk assessment. In the past, a decision about a new customer’s ability to pay might be based on something like higher income variables. But when DXM put all the data back into the data warehouse and built a machine-learning model, it found that people’s ability to pay became better with age, not income. The more mature they were, the more diligent they were at keeping up payments. DXM tried other variables, such as geography and gender, but the machine-learning model kept coming back, and telling it that the correlation of ability to pay with age was the strongest overarching variable.

Then, finally, because DXM possessed a large amount of geospatial data on how far prospects were traveling to make a purchase, it could tell the telecom retailer how to maximize its marketing trade area beyond its anecdotal boundaries. This wasn’t an online scenario —you had to actually go into a store, pick out a phone, and make your purchase—but DXM found that 62% of consumers traveled more than six miles to make their purchase, whereas the client had been capping its trade area at five miles.

Overall, DXM was able to reduce customer acquisition costs by 52%, deliver analytics projects to clients 70% faster, and grow revenue 25% in six months by using machine learning attached to its data warehouse.

One of the bigger revelations in the digital world was that DXM also found that up to 88% of consumers who end up buying a product or services will never touch the ad, never click it, never share it, never like it. They do respond to the ad by eventually purchasing the product advertised, but by coming in and converting on their own terms.

"That's been big for our clients—to get that much insight into folks who are actually buying from them," says Owens. "That has been some of the most important data coming out of the machine-learning initiative, that we can actually see the correlation of all these logs—and that impressions do matter if they're delivered to the right audience, and if you're able to track the journey to conversion. You can't do that with so many of these walled gardens that refuse to give you the transparency for your marketing spend."

He adds, "Having access to this type of machine-learning capability, we just get quicker to seeing these trends, these behavioral intents, and that we can get to market, and actually get conversions going pretty quickly for our clients."

Looking Forward: Better Hardware and Automation

The goal of machine learning is to find patterns in large amounts of data and to use insights into those patterns to make predictions about future events or scenarios. It is a particularly hot topic these days, and most organizations either have started or plan to start machine-learning initiatives in the near future.

The future of machine learning lies not just in better algorithms but also in hybrid data management and machine-learning platforms that are able to speed machine learning by moving algorithms to the data, and not the other way around. It also lies in advanced hardware, which is rapidly evolving. We are seeing faster and more sophisticated generations of GPUs that will allow you to train machine-learning models very quickly. This is another reason to go to the cloud, because if you make an investment in on-premises hardware today, it will take you longer to move to the next generation of innovation tomorrow.

We will also see a lot more automation, on both the model training and monitoring sides. On the training side, AutoML is rapidly evolving, which will reduce the work overhead for the data scientists and democratize machine learning and data science even further.

On the monitoring side, we're entering an age where data science platforms and cloud providers are beginning to focus more on that aspect of the machine-learning life cycle.

Much exciting work is being done in machine learning across a broad range of industries. In particular, new cost-effective and infinitely scalable cloud-based platforms are emerging that streamline businesses' capabilities to build machine-learning models for data analytics, data discovery, and data visualizations in a way that promotes collaboration among data scientists, data engineers, business analysts, and business users for best business outcomes. The future looks bright for early adopters of machine learning who are getting ahead of the curve as well as the competition.

About the Author

Alice LaPlante is an award-winning writer, editor, and teacher of writing, both fiction and nonfiction. A Wallace Stegner Fellow and Jones Lecturer at Stanford University, Alice taught creative writing at both Stanford and in San Francisco State's MFA program for more than 20 years. A *New York Times* best-selling author, Alice has published four novels and five nonfiction books and also has edited best-selling books for many other writers of fiction and nonfiction. She regularly consults with Silicon Valley firms such as Google, Salesforce, HP, and Cisco on their content marketing strategies. Alice lives with her family in Palo Alto, California, and Mallorca, Spain.