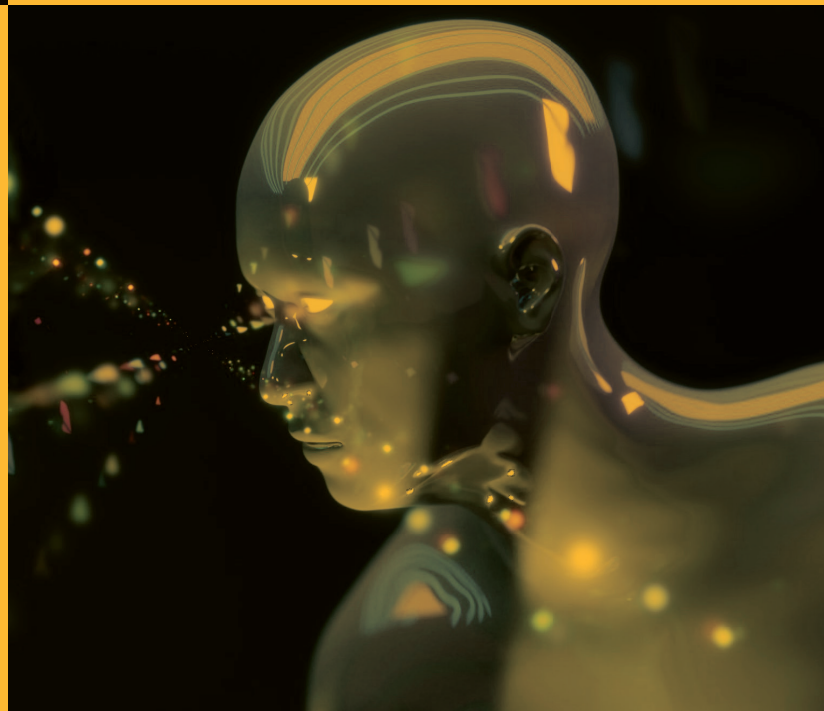


ActiveState®



EXECUTIVE GUIDE TO MACHINE LEARNING

WHAT YOU
NEED TO
KNOW BEFORE
YOU START



"Artificial intelligence will, on average, boost rates of profitability by 38% and provide an economic boost of \$14 trillion in additional gross value by 2035"

Accenture





CHAPTER 1

WHY MACHINE LEARNING?

Once a generation (or so), a game-changing innovation comes along that disrupts businesses across all industries, and changes not only the way we work, but the way we live.

This generation will be defined by the rise of Artificial Intelligence (AI).

AI is driving the creation of systems that can not only learn and adapt, but may at some point even act autonomously. But for now, AI is primarily being used to augment human activity via Machine Learning (ML); where ML is focused on giving machines ac-

cess to data and letting them learn for themselves.

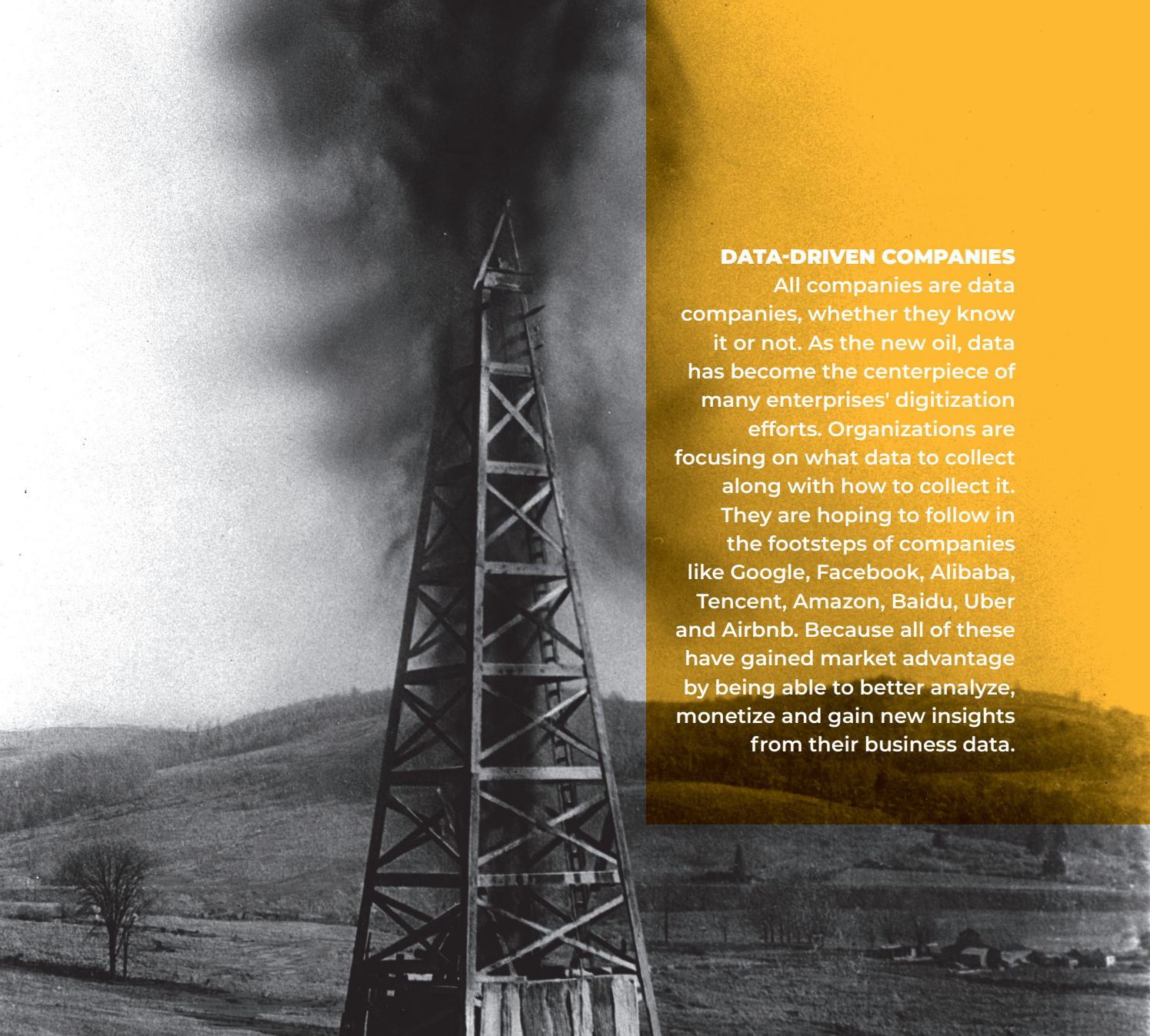
Based on data collected in the enterprise, ML is being used today to enhance decision making, disrupt traditional business models, and redefine the customer experience.

Over the next few years, every innovative application and service will incorporate some level of ML. This trend dovetails nicely with traditional industries that are currently undergoing a digital transformation to become more data driven.

ML in Postal Services

USPS has previewed their "Smart Blue Box" project, which attaches a voice-based interface to any regular mailbox, enabling a mini post office on every street corner. The voice-activated mailbox uses AI to answer consumers' questions, such as "How much will it cost to ship this package?" The package is weighed on the spot and shipped off in half the time, according to the USPS ¹.

¹ <https://youtu.be/qGOeznUuy-Y>



DATA-DRIVEN COMPANIES

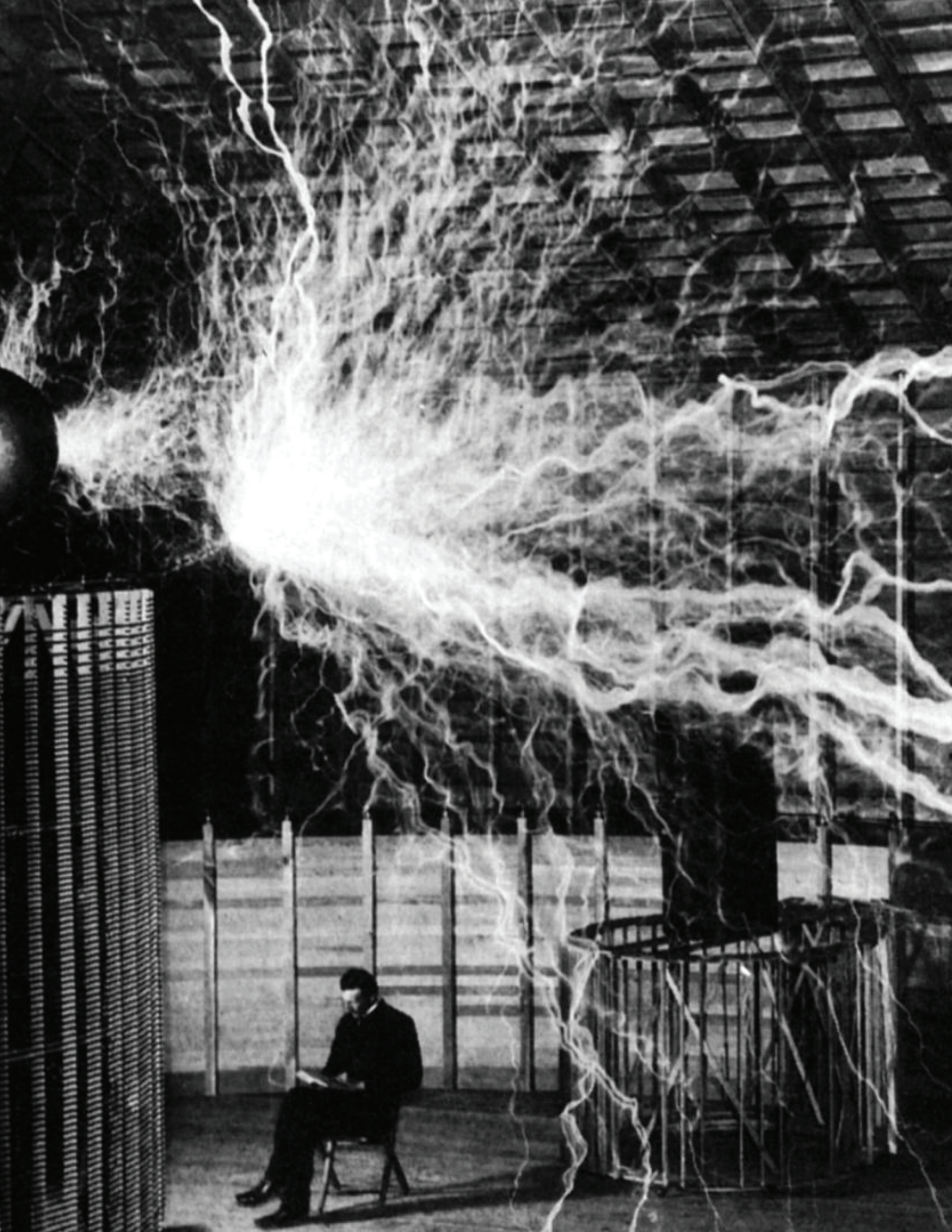
All companies are data companies, whether they know it or not. As the new oil, data has become the centerpiece of many enterprises' digitization efforts. Organizations are focusing on what data to collect along with how to collect it. They are hoping to follow in the footsteps of companies like Google, Facebook, Alibaba, Tencent, Amazon, Baidu, Uber and Airbnb. Because all of these have gained market advantage by being able to better analyze, monetize and gain new insights from their business data.

But data volume and the increase of number of variables collected makes for another set of challenges. It's either impossible or impractical to explore

every possible pattern and determine whether the findings are relevant and actionable with current visual-based data discovery tools. Enter ML.

ML in Finance

A JPMorgan Chase program, called Contract Intelligence (aka COIN), does the mind-numbing job of interpreting 12,000 commercial-loan agreements in seconds compared to the 360,000 hours of work previously required each year by lawyers and loan officers. The result is not only speed, but also a decrease in loan servicing errors.



THE ML-DRIVEN COMPANY

If data is the new oil, then ML is the new electricity.

The next-generation of data insight tools now feature ML at their core.

In this context, ML automation augments human intelligence and contextual awareness. It removes human bias from the equation and identifies patterns not obvious to the naked eye.

The output will optimize the decisions and actions of all employees from HR to finance to sales, marketing, customer service, procurement and asset management, no matter the industry.

SO HOW DO YOU GET STARTED?

The good news is that the Big Data initiative your company started a dec-

ade ago is finally going to pay off. The bad news is that the ML experts and data scientists you need to make that happen are scarce resources.

If you find yourself in the planning stages for ML but not ready to implement, you're not alone. Gartner estimates that only about 12% of their customers have either implemented data science and ML, or are implementing it right now ².

This guide is intended to help clarify your options as you evolve toward an ML-driven company.

ML in Transportation

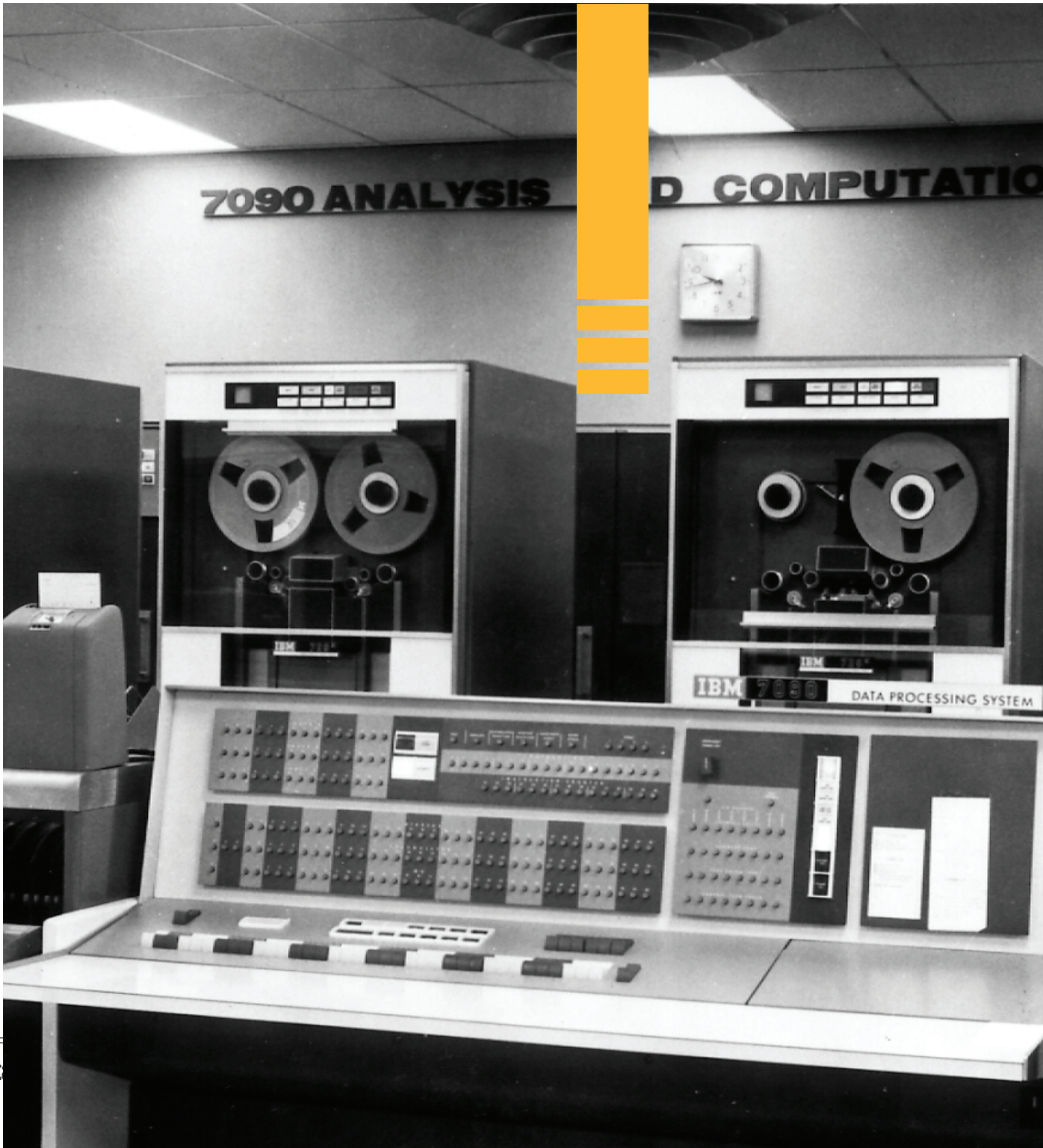
Lufthansa Industry Solutions is applying ML to reduce flight delays by predicting everything from whether a passenger will show up for the flight to optimizing refueling times to predicting the arrival time of an aircraft at the moment of departure. ³

² Market Guide for Data Science and Machine Learning Service Providers, 31 October 2017

³ <http://dataconomy.com/2017/10/structuring-it-at-lufthansa/>

"Healthcare, financial services, and professional services are seeing the greatest increase in their profit margins as a result of AI adoption."

McKinsey Global Institute, Artificial Intelligence, The Next Digital Frontier





CHAPTER 2

ML STARTS WITH BIG DATA

The most logical starting point in any ML project is the data. ML thrives on data – the more, the better, since machine learning models typically require a substantial data set in order to train them. That’s good news for enterprises that have spent the past decade setting up and operationalizing a Big Data initiative. If you’re starting from scratch you may want to consider jumpstarting your skills by leveraging public data sources such as those found through Data.gov initiatives or general data fit for ML.

BUT KEEP IN MIND TWO THINGS

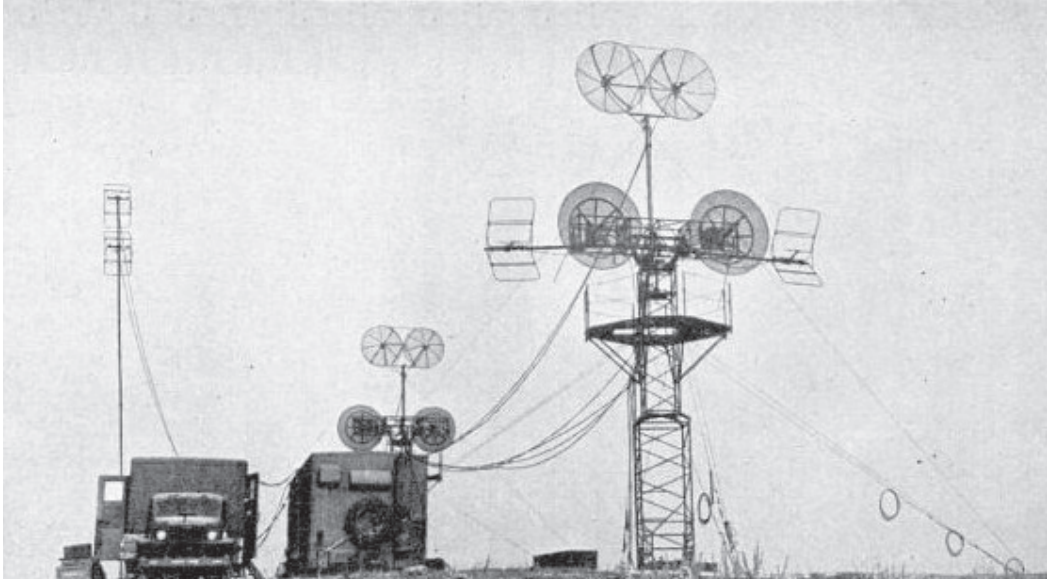
- ▶ Machine learning models are only as good as the data they’re trained on; and
- ▶ An organization’s ability to convert data into value is largely correlated with the maturity of their data infrastructure.

ML in Healthcare

Sophia Genetics uses its AI called “SOPHiA” to analyze each of their patient’s genomic profiles. The output allows them to diagnose illnesses and better understand each patient’s unique health requirements. By combining AI with DNA, companies like Sophia Genetics are ushering in a new era of personalized medicine that offers better outcomes than standard treatment.¹

¹ <https://siliconangle.com/blog/2017/09/13/sophia-genetics-raises-30m-ai-reads-dna-diagnose-illnesses/>





DATA ENGINEER & DATA SCIENTIST: TWO KEY ROLES TO NAVIGATE THE BIG DATA LANDSCAPE

Data Engineers have familiarity with BI, data warehousing and software engineering needed to create the kind of big data system that can feed your ML project. Without a solid, foundational data infrastructure, every activity related to your data science initiative will become either too expensive, not scalable or both. You never want your data science team to query straight from the production database, which can cause

delays at best, and outages at worst.

Recently, the market has shifted toward a connect rather than a collect approach. As a result, Data Engineers have also become responsible for the creation and operation of big data distributed systems centered around distributed computing technologies like Hadoop and Spark. These technologies are now able to process up to petabytes of data – a feat that would have been impossible just a few years ago.

ML in Stock Trading

At Goldman Sachs, complex trading algorithms replaced trades where the market price of what's being sold was easy to determine. Now, more complex areas of trading like currencies and credit, which go through networks of traders, have algorithms designed to emulate what a human trader would do. As a result, the work of 4 traders can be replaced by 1 computer engineer.³



DATA SCIENTISTS, UNICORNS & OTHER SO-CALLED FANTASIES

Data scientists, on the other hand, tend to be not only math and statistics focused, but also highly skilled at finding insights in data. Unfortunately, demand currently outstrips supply for these individuals, especially for those that also have either programming and/or domain skills.

If you're lucky enough to hire these kinds of unicorns, you can centralize your data science and ML efforts in order to go deep and get the biggest bang for your buck. If not, it may make more sense to go wide, by

nurturing the skills you already have across your organization. In fact, by growing data science skills in every department, you can apply machine learning closer to the domain problems. And therefore scale ML across the organization. While this approach may sound like a pipe dream, as we'll see in Chapter 3, there have been recent advances in a number of tools and platforms that make ML more accessible than ever to a wide range of IT professionals.

ML in Insurance

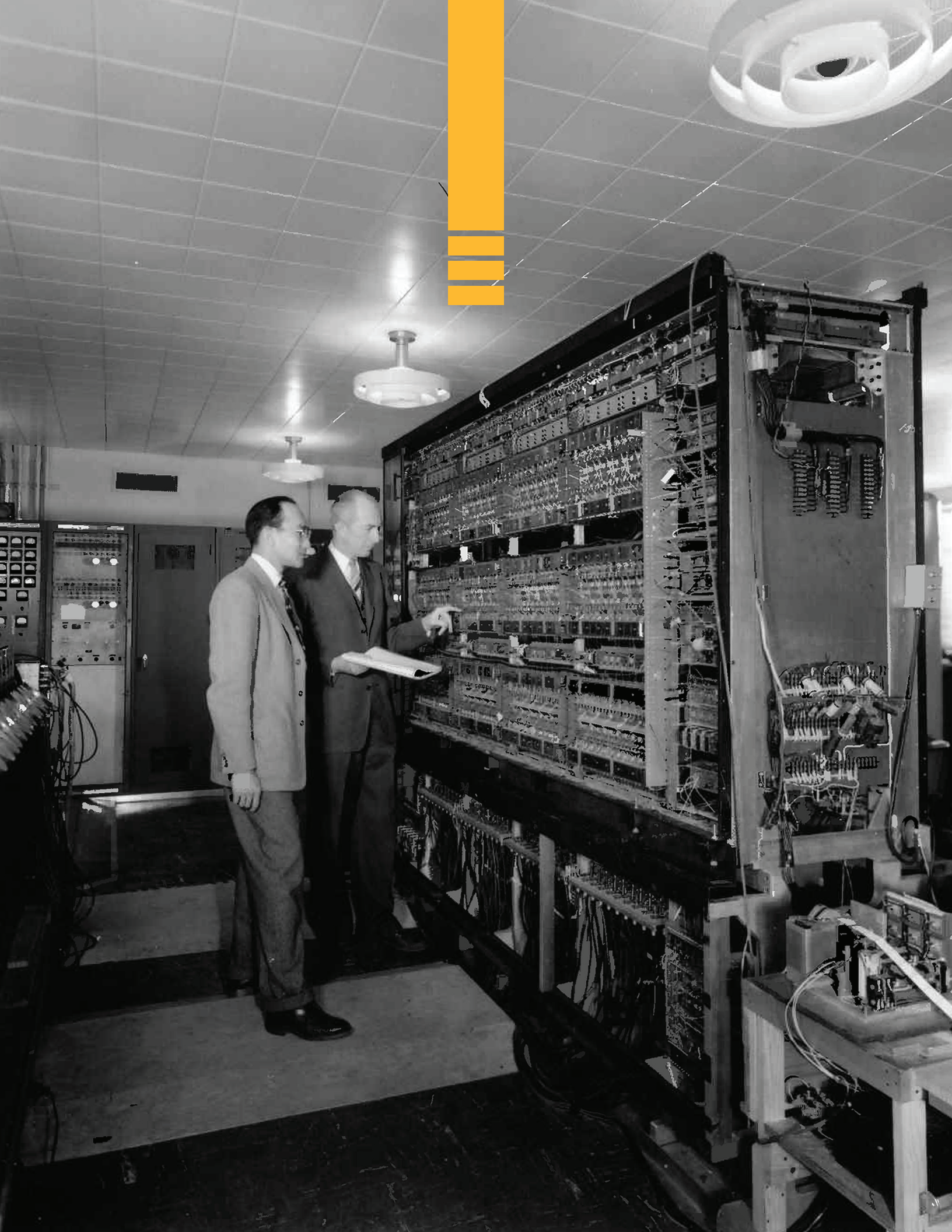
US-based startup Lapetus has designed a machine learning program that uses "selfies" to underwrite policies. This eliminates the manual process of applying actuarial models to customers. Instead of waiting days for a decision, the ML algorithm is able to examine gender, rate of ageing, and body mass index from a single photo in order to instantly predict life expectancy.

³ <https://www.technologyreview.com/s/603431/as-goldman-embraces-automation-even-the-masters-of-the-universe-are-threatened/>

⁴ <https://www.redpixie.com/blog/machine-learning#227423290-tw%231503376043986>

"The global cognitive computing market is expected to reach \$12.5 billion in 2019, up from 2.5 billion in 2014, at a CAGR of 38%."

Research and Markets



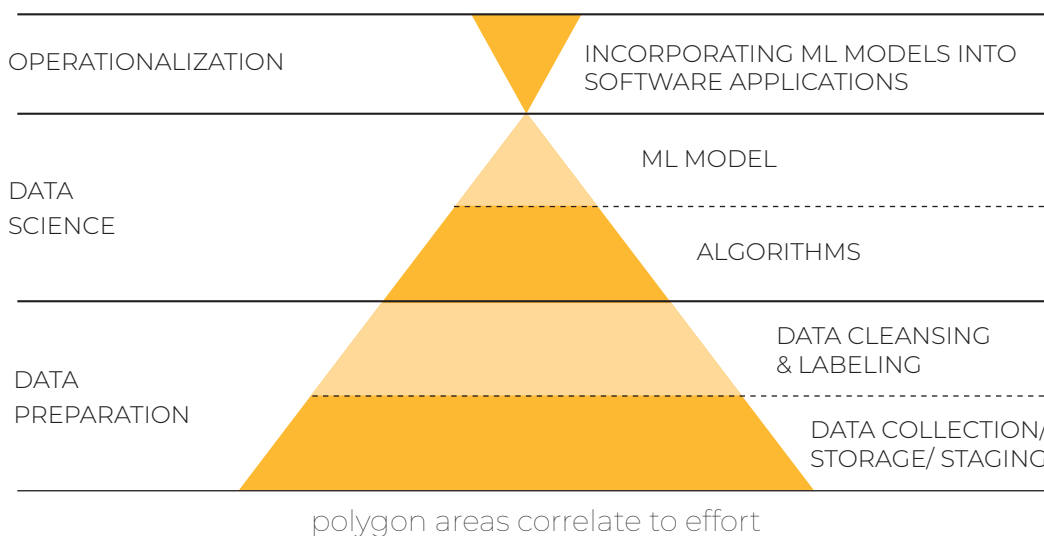
CHAPTER 3

COMMERCIAL VS OPEN SOURCE

A comprehensive ML solution provides features and functionality that encompass data preparation (e.g., data cleansing, labeling, etc.), data science (e.g., algorithms and model building), and operationalization (i.e., incorporating ML models into software applications). The data science effort is very compute-intensive. By comparison, the data preparation work is labor-intensive. As a result, a significant proportion of your organization's work will be focused on preparing the data: organizing it into appropriate domains, cleansing it of bad data, finding sub-

ject matter experts to label the data correctly, and creating the data pipelines that will feed your model.

But at the end of the day, ML is all about math, specifically statistics and algorithms, two areas that have been addressed by software solutions for decades. What's new is the price tag. Previously these kinds of solutions used to be prohibitively expensive to purchase, deploy and maintain, but today organizations can choose from either a growing number of affordable commercial solutions or freely downloadable open source options.





The advantages and drawbacks of open source versus commercial solutions for general software development hold true for solutions in the ML space as well, namely:

- ▶ Open source does not require a license fee, and therefore attracts a sizeable community
- ▶ Large open source communities provide timely bug fixes, as well as answers to common questions
- ▶ A large community also makes it easier to recruit technical resources

Commercial ML solutions, on the other hand, provide:

- ▶ SLA-backed support and maintenance
- ▶ Time savings: pre-defined and pre-integrated functionality save time over cobbling together a set of open source modules from different authors

ML in Online Services

Amazon's recommendation engine is perhaps the most famous example of ML being used to directly influence online user purchase decisions with a direct impact to top line revenues. Another example is how Netflix leverages ML to reduce customer churn. Netflix examines user viewing preferences to recommend similar movies which results in an estimated ROI of more than \$1B per year through increased customer retention.¹

¹ <https://www.redpixie.com/blog/examples-of-machine-learning>

As a general rule of thumb, open source solutions provide the greatest agility and control, and are best suited for organizations that have unique data or for whom analytics provides a key, unique differentiator. However, an open source approach will require strong, in-house ML skills. By comparison, commercial solutions allow less skilled organizations to get started right away by providing access to pre-built frameworks featuring popular algorithms that can decrease time to solution. However, commercial off-the-shelf solutions may prove limiting as you attempt to create white space with your competitors who will also be jumping on the ML bandwagon.

While there are many ML open source solutions available today, the two most popular are based on Python and R. Python has recently overtaken R as the most commonly used solution for ML. Whereas R is still popular among statisticians and general data science applications, Python now incorporates the bulk of all ML libraries, including Google's

TensorFlow and Microsoft's Cognitive Toolkit. These toolkits streamline the adding of more layers to your neural network, where each layer adds the ability to examine a more complex detail of the problem you're trying to solve. Python also includes one of the most popular data preparation libraries currently available (Pandas), and offers a wide range of web and application development packages for operationalizing ML models.

One of the best ways to get started with open source is via the cloud, which can easily scale to handle high data volumes, but may not be appropriate for organizations with data sensitivity issues. Amazon Web Services, Google Cloud Platform and Microsoft Azure can all provide you with a pre-installed set of ML open source toolkits as an "out of the box" virtual machine. Moreover, many of these "ML as a Service" (MLaaS) offerings build in the ability to operationalize trained models directly as APIs, in effect, building in the expertise needed to get successful results.

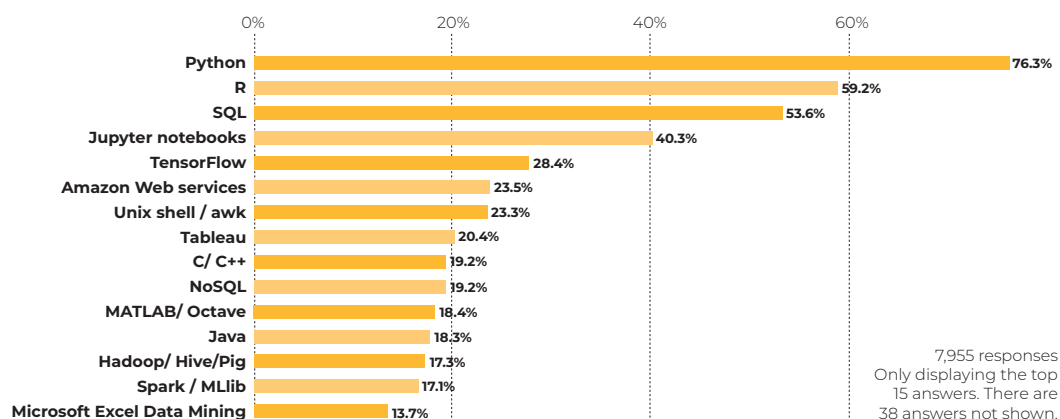
ML in Sales & Marketing

Salesforce.com and Hubspot use ML to help distinguish signal from noise when it comes to examining contact, lead and customer interactions across multiple touchpoints. The result is a deeper understanding of the prospect, which can lead to highly personalized messages, improved lead scoring and more, better qualified opportunities. ²

² <https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications>

WHAT TOOLS ARE USED AT WORK?

Python was the most commonly used data analysis tool across employed data scientists overall, but more Statisticians are still loyal to R.



The State of Data Science & Machine Learning ³

While there is good consensus around Python and R in the open source space, the commercial space is much more fragmented with a huge variety of products, many based on or incorporating open source components. However, there are some acknowledged leaders and innovators, including:

SAS VISUAL ANALYTICS BY SAS INSTITUTE

Provides a comprehensive portfolio of data science solutions that extend from data preparation, analysis and visualization to model building and deployment. SAS's strength lies in being able to process huge data sets (>1B records) and present results in an easy-to-digest, web-based dashboard. However, manipulating data sets in any custom way requires

learning SAS' proprietary programming language. SAS is widely considered the leader in the commercial ML market, both by number of clients and revenue, despite oft-cited high licensing costs.

MATLAB BY MATHWORKS

Is a proprietary programming language widely adopted by scientists for numerical analysis, quants for financial engineering, and academics for statistical analysis. As an ML toolbox, it also offers a selection of classic algorithms, pre-trained models, and applications/interfaces to C, .Net and Java to help operationalize ML models. Much like the SAS solution, MATLAB has been in the market for decades, and has a large install base that cite licensing costs and vendor lock-in as primary concerns.

³ <https://www.kaggle.com/surveys/2017>

H2O.AI

Is a market innovator trying to make ML accessible to businesses that don't necessarily have ML expertise. Their solution makes it easy to apply standard algorithms against your dataset in order to create models for discovering patterns in claims processing, credit scoring, customer churn, or fraud detection data, as well as for discovering general operational or predictive insights. In contrast to SAS and MATLAB, H2O.ai simplifies ML ramp up, but may have less customizability options to deal with more complex problems.

When selecting any software vendor, it's tempting to opt for a feature-rich, proven commercial solution. However, mismatches between your organization's capabilities and the vendor's strategy can derail your ML efforts. Ensure that your organization's risk tolerance and – more importantly – culture are aligned with your vendor, since you will need a partner who can grow and adapt at the rapid pace with which the ML market is moving.



ML in Aerospace

A fighter jet AI developed at the University of Cincinnati called ALPHA defeated a US Air Force trainer, as well as all other AI's available at the Air Force Research Lab – all while running on a Raspberry Pi. What distinguishes ALPHA is its fuzzy tree decision system that calculates movements and strategies 250 times faster than its opponents can blink. ⁴

⁴ http://magazine.uc.edu/editors_picks/recent_features/alpha.html

"Tech giants including Baidu and Google spent between \$20B to \$30B on AI in 2016, with 90% of this spent on R&D and deployment, and 10% on AI acquisitions."

McKinsey Global Institute, Artificial Intelligence, The Next Digital Frontier





CHAPTER 4

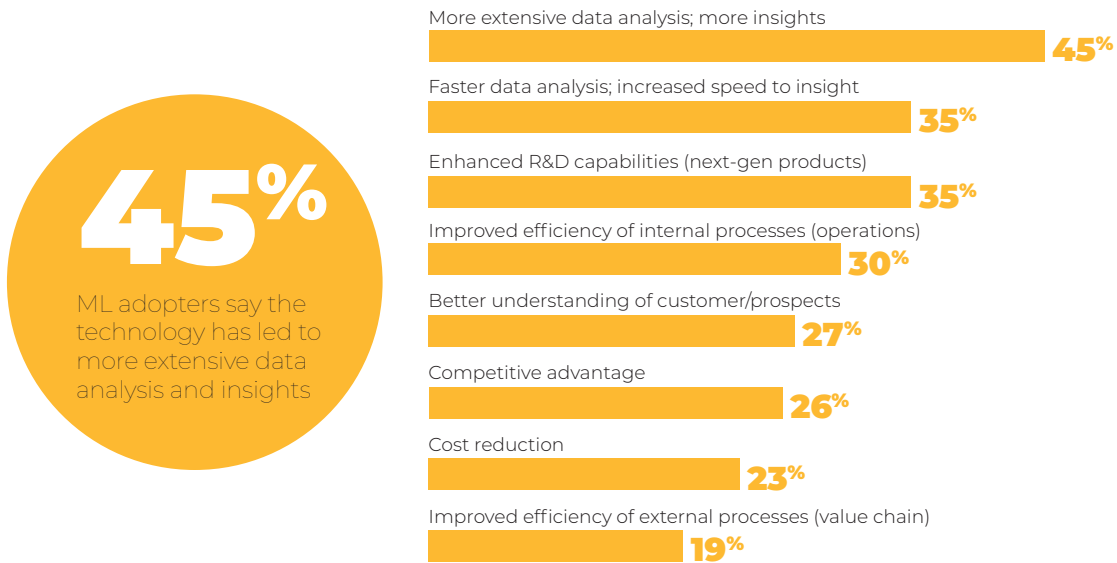
BECOMING AN ML-DRIVEN COMPANY

By now, you've discovered why you need ML, where to start (with the data), and how to select an ML solution. Now all you need to know is what to work on.

Start by assessing which business outcomes would deliver significant ROI to the business, but have traditionally been hard-to-solve or are personnel intensive. For example:

- ▶ ML is well suited to analyzing complex datasets, such as those associated with making predictions or forecasting.
- ▶ Avoid the trap of "ML project as a fishing expedition." Use traditional analytics to help you first locate data anomalies or datasets that feature multiple interdependent variables and noisy data.
- ▶ ML routines can be used to replace personnel involved in recurring activities, such as CSRs, retail purchasing agents, etc.
- ▶ ML can help mitigate risky business processes such as decision making in a loan approval process. Any process that would otherwise require extensive training of personnel to yield results within an acceptable error rate is a prime candidate.
- ▶ ML can improve efficiency in decision-making processes with clear, quantifiable inputs and outputs, such as stock trading.
- ▶ ML can learn to do mundane tasks such as classification of photos, video, and other traditionally personnel-intensive identification tasks.

Q10. If your organization is currently using ML, what have you actually gained?*



*Multiple responses allowed. All percentages have been rounded

You should also learn what you can from successful case studies in your industry.

ML and the Traveling Salesman Problem

Mathematicians have been trying to find the most efficient route between cities for salesmen before they return home since 1930. Wise Systems, an MIT spinoff, is applying this concept to delivery truck route optimization using data like speed, GPS location, traffic, weather, destination, and even custom data (e.g. parking availability) that drivers enter via a mobile app. But Wise adds an ML twist: its algorithms learn from each day's data recorded for each customer in order to incrementally improving routes over time. Anheuser-Busch has reduced the miles traveled per stop by 4%, which translates into fuel savings, and lower wear and tear on trucks.

¹ <http://fortune.com/2018/02/06/machine-learning-delivery-driver/>



MAKING THE CULTURAL SHIFT

Once you have a clearly defined mandate, place your data scientists in the business unit that will benefit. Being close to the subject matter experts means ML outcomes will better align with that unit's goals.

As a result, frontline managers can be armed with insights that allow them to start thinking more critically, sharing more data, and learning to make more decisions on their own in order to create value for customers and the business by creating new services or adding value to existing offerings.

The goal here is not only to justify the cost and complexity of your investment in ML, but more importantly, to normalize it as a means of doing business across your enterprise. Change is hard for any organization. Do not underestimate the number of silos that will need to be broken down to become an ML-driven company.

Simply buying ML software and implementing it as a new, enabling technology will not be enough to succeed. Rather, think of ML as a tool to help you craft and implement a strategic vision that can only come about as a result of a number of behavioral changes and tactical initiatives, including:

- 1.** Creating consensus and executive buy-in. This may require the creation of a proof of concept before the benefits can be made apparent.
- 2.** Preparing your data. Collecting, cleaning, labelling and staging your data may require a sizable time investment depending on the maturity of your big data efforts.

The benefit here is that ML works well with messy, incomplete and noisy data sets – the exact kind of data that executive decision makers typically distrust.

- 3.** Finding the right algorithms. Be prepared for a lot of trial and error here as you test multiple algorithms against your data. The good news is that the computing resources required to discover a successful ML model are faster and cheaper than ever.

The bad news is that ML models are “black boxes,” meaning it’s not possible to explain with confidence how they make their decisions. Some industries like housing, finance and healthcare have regulations requiring you to give explanations for decisions. Document not only how your ML model delivers accurate predictions, but also the rationale for those predictions.

4. Operationalizing your models. Many organizations struggle with this last mile that requires incorporating your ML model into a software application.

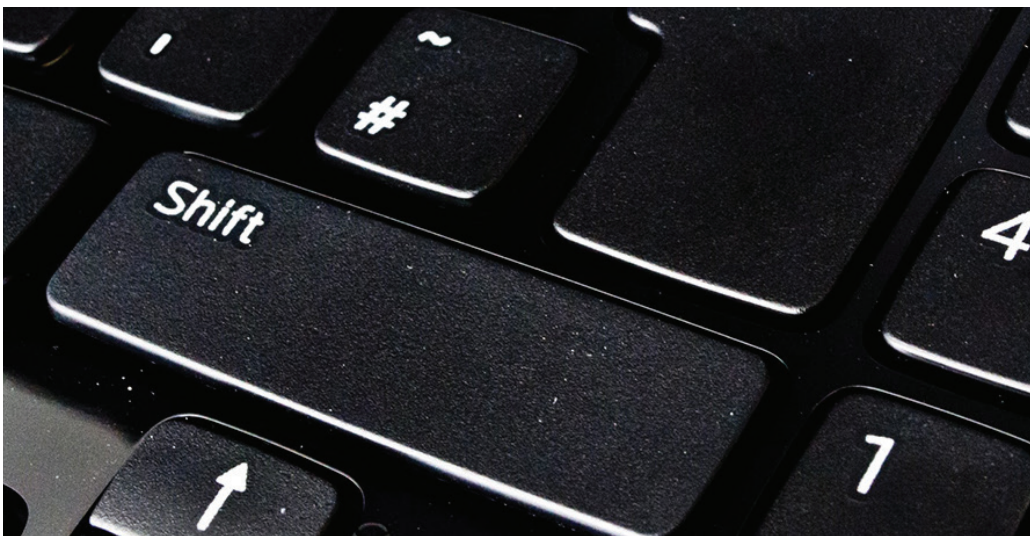
Ensure your data science and software development teams can work closely together, ideally sharing the same tools and/or platforms. In this way, the process of handing off ML models will be more of a collaborative effort than a “drop and run” event as the data science team moves on to the next problem.

5. Be prepared to retrain your models as your business evolves. Businesses (and the data you collect) change over time, which means the performance of ML models will also degrade over time, yielding less accurate results.

ML in the Credit Card Industry

Fraud detection can sometimes seem magical, but is typically based on data analysis techniques that previously required significant human involvement. Today, the industry trains machine learning algorithms on the data compiled from each consumer’s daily credit card usage over a number of years. Now, when you swipe your card the amount, time stamp, merchant identifier and other details go to the card issuer where the ML model evaluates whether it might be a fraudulent transaction in real time.

³ <https://www.kaggle.com/surveys/2017>

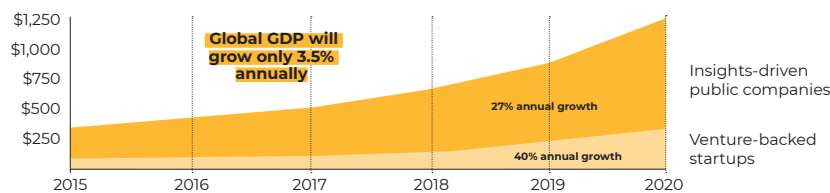


SUMMARY

ML represents both a challenge and an opportunity for every organization. The key take-away is: you need to invest in ML today to avoid obsolescence. ML will create new services, products and offerings, transform industries and level playing fields. There are few (if any) industries that will not be disrupted by a technology that endows machines with human reasoning capabilities backed by near-limitless computing power.

However, ML remains nascent: no one organization has achieved dominance. While tech giants like IBM, Microsoft, Amazon and others are investing heavily today, they are likely to be outstripped by the ML startups who are receiving record-breaking venture capital funding. Organizations that can make the fundamental changes needed to integrate ML into their strategies and day to day operations still have a chance to win.

Revenue forecast of insights-driven business (\$ billions)



Note: The data point public companies in 2015 is actual revenue; all other data points shown are estimates or projected figures.

Source: Economic Intelligence Unit, Morningstar, and PitchBook Data

ML-Driven Enterprises Will “Steal” \$1.2 Trillion Annually By 2020 from Non-ML-driven Companies ³

ML in Manufacturing

GE’s Brilliant Manufacturing Suite links design, engineering, manufacturing, supply chain, distribution and services into one globally scalable, intelligent system powered by Predix, GE’s industrial internet of things platform. Predix uses sensors to capture every step of every process, and uses ML to spot potential problems and solutions. According to GE, “Brilliant Factories” like their wind generator factory in Vietnam increased productivity by 5%, and its jet engine factory in Muskegon had a 25% better on-time delivery rate.⁴

³ https://go.forrester.com/wp-content/uploads/Forrester_Predictions_2017_-_Artificial_Intelligence_Will_Drive_The_Insights_Revolution.pdf

⁴ <https://www.techemergence.com/machine-learning-in-manufacturing/>

"By 2020, 57% of business buyers will depend on companies to know what they need before they ask for anything. This means having solid prediction capabilities with your AI will be the key to keeping your customers. "

(Salesforce)

Fast track your ML projects with
ActiveState's Python packages and tools.

Contact us to find out why 97% of
Fortune 1000 use our software:

solutions@activestate.com

ActiveState[®]

website: www.activestate.com | Toll-free in NA: 1.866.631.4581 | email: solutions@activestate.com

ABOUT ACTIVESTATE

ActiveState is a leader in providing commercial level open source language distributions. It provides commercial versions of Python, Tcl, Perl, Ruby and Go. More than two million developers and 97% of Fortune 1000 companies use ActiveState open source language builds including CA, Cisco, Pepsi, Lockheed Martin and NASA. To learn more, visit activestate.com