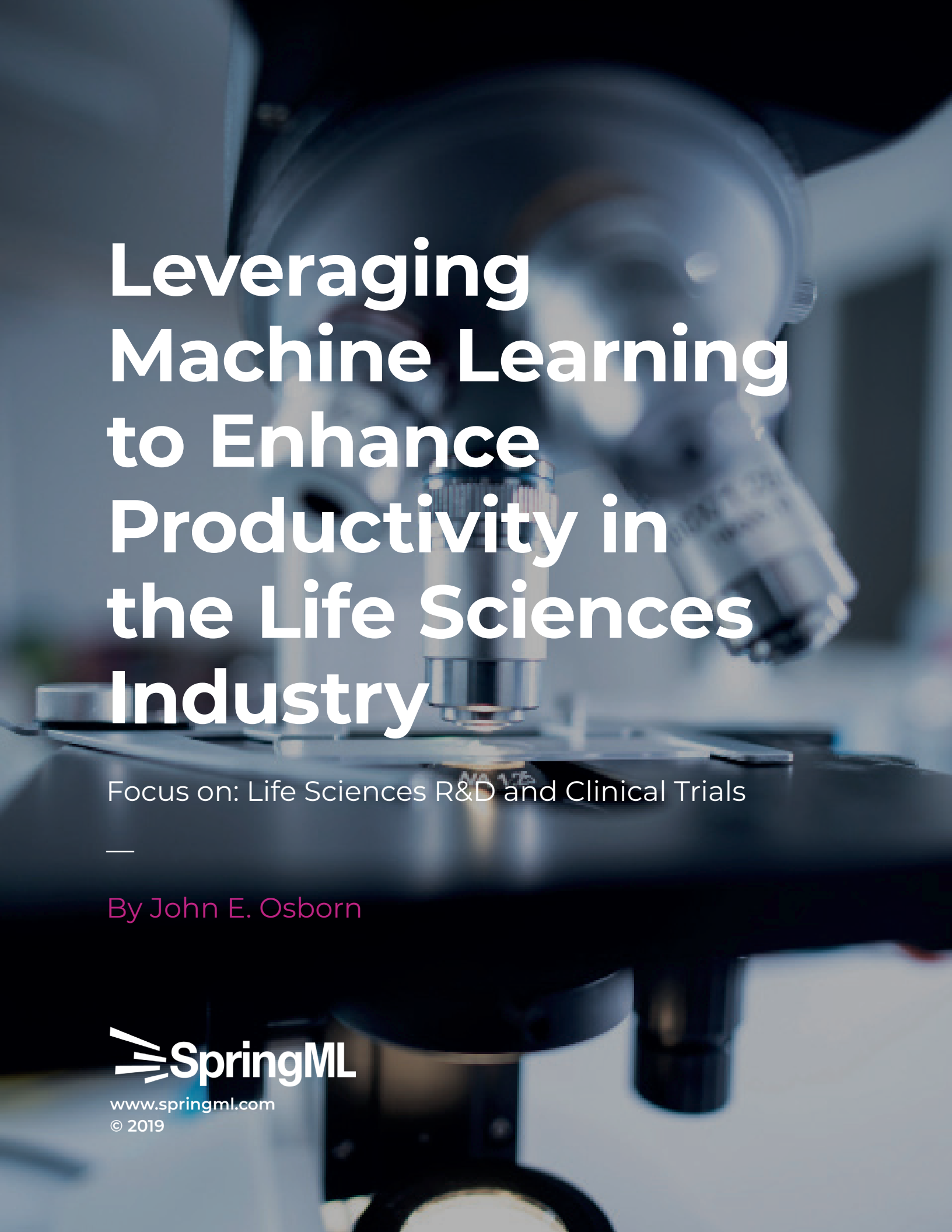


Leveraging Machine Learning to Enhance Productivity in the Life Sciences Industry

Focus on: Life Sciences R&D and Clinical Trials

—

By John E. Osborn

SpringML

www.springml.com

Table of Contents

About the Author

John E. Osborn
is an affiliate professor at the University of Washington, Seattle, a regular contributor to Forbes.com, a former life sciences company executive, and a former member of the U.S. Advisory Commission on Public Diplomacy. John also has served on corporate boards and advises life sciences and healthcare companies. He is working with SpringML on a project that included the preparation of this paper.

Drug development, precision medicine, and the case for Machine Learning

The time required to identify a compound, demonstrate that it is both safe and effective for a specified condition, and obtain the requisite regulatory approvals to bring a new drug to market is staggering. Even after a molecule or biologic is identified as having potential therapeutic value, drug development work typically takes ten years or more to complete, with human clinical trials taking up 6 to 7 of those years.[1] While industry observers and critics debate the assumptions and methods used to calculate the average cost of drug development, it is undeniably expensive. Estimates range from $757[2] million up to $2.6[3] billion.

These costs would be prohibitive if not for the promise of revenues that offer a meaningful ROI of several-fold or more upon approval and successful commercialization. Yet there is much risk associated with this endeavor, as the high cost of drug development is accompanied by correspondingly low rates of success. Even with firms using mechanized screening tools to rapidly assess the viability of thousands of potential compounds, PhRMA puts the *"probability of clinical success (the likelihood that a drug entering clinical testing will eventually be approved) at less than 12%."*[4]

To complicate matters further, the days of the so-called "blockbuster" drug appear to be waning. The promise of multi-billion-dollar payoffs for therapies with broad indications that treat a common disease or illness faced by masses of patients is becoming an exceedingly rare occurrence. Instead, drug companies are focused on identifying novel targets that address a rare disease or a specific type of cancer; this is the essence of what we have come to call precision, or personalized, medicine.

The move to precision medicine is linked to therapies that target biological markers of disease associated with specific genomic subtypes. While this development points to a new age of drug discovery for life sciences firms, it also calls for a more agile development process capable of efficiently sorting massive volumes of data to gain scientific and clinical insights.

The importance of data stems from a number of developments. A few of these include the success of the human genome project, the digitalization of patient health records, and advances in cloud computing. Today, firms can access data sets that are unprecedented not only in their volume but in their variety.

In such a data-intensive environment, the challenge is to access and evaluate patient health data efficiently. In a recent statement, Dr. Scott Gottlieb, Former FDA Commissioner, puts it simply:

"Without a more agile clinical research enterprise capable of testing more therapies or combinations of therapies against an expanding array of targets more efficiently and at lower total cost, important therapeutic opportunities may be delayed or discarded because we can't afford to run trials needed to validate them."[5]

With costs soaring and business models predicated on continued innovation, life sciences firms are searching for greater efficiency and agility. One of the most promising ways forward lies in harnessing the power and potential of machine learning (ML) tools.

Machine learning is the process of detecting patterns in large data sets using algorithms that parse the data to reveal connections and develop insights from the data. Instead of following a set of hard-coded rules or routines, the algorithm exists within a framework, evaluates the data, and derives a set of rules for measuring the significance of certain factors without human intervention.

Algorithms are initially "trained" using data sets that provide clear examples or patterns upon which they can build. Probabilistic in nature, the insights generated from ML become more dependable over time as the algorithm churns through larger and larger data sets and improves its performance in an adaptive fashion.

The advantage for life sciences firms is that ML provides a way to unlock the hidden value of big data in ways that no individual or team could possibly hope to do on their own. The insights generated can increase the odds of identifying effective targets, help optimize product portfolios, improve clinical trial success rates, and mitigate risk by ensuring quality in the supply chain.

In this paper, we consider several ways in which life sciences firms can use machine learning to favorably impact R&D productivity, clinical efficacy, and supply chain efficiency.

Drug discovery: Mining for diamonds in the data

Machine learning is fast becoming an indispensable tool for life science companies struggling to identify and develop innovative new therapies in the age of precision medicine. As the leading edge of drug discovery and development has moved to large molecules and more targeted therapies that follow unique biological pathways, the volume of data generated in pursuit of these therapies can be leveraged by the increased use of machine-aided insight.

As the leading edge of drug discovery and development has moved to large molecules and more targeted therapies that follow unique biological pathways, the volume of data generated in pursuit of these therapies can be leveraged by the increased use of machine-aided insight.

Before applying ML, it is important to consolidate all relevant data within a secure cloud-based storage system. While some companies have substantial data storage capabilities on their own premises, data consolidation on premise likely will not support the multi-functional (and perhaps even multi-entity) access necessary to support widely dispersed collaborative teams working across geographies. This is true for all use cases explored in this paper.

Once the required data are consolidated for ready access by researchers, R&D teams can then consider applying ML to enhance understanding and leverage their work. In fact, many life sciences firms have come to believe that, while we are still in the early stages of applying ML to solve real-world problems in human biology and disease, it has the potential to help firms discover those rare “diamonds” hidden in their data that would otherwise be impossible using traditional tools.

One area of focus for ML is product line extensions. Can an existing therapy work for new patient populations with slight modifications? Or can a previously unsuccessful molecule be repurposed for a subset of the patient population with a particular set of biomarkers – as in the example of Merck’s Keytruda® (pembrolizumab), which was the first treatment to be approved by the FDA only for patients with cancer and a specific genetic feature?

As treatments become more targeted, they are also becoming more complex. ML can help scientists move beyond binary modes of analysis to consider a quantitative blend of multi-factorial data points. Increasingly, the medicines of the future will involve composite sets of biomarkers and equally complex mixes of existing and new drugs to treat the disease or condition. ML is seen as indispensable for dealing with this level of complexity.

Therapies targeted at the genomic level are getting attention from firms partly because the genomic data is available in enough volume and density to make ML an effective tool for analysis and discovery. For example, using next-generation sequencing (NGS), a high-throughput approach to rapidly sequencing DNA, researchers can identify patterns in the genomic data to distinguish different disease phenotypes and identify relevant biomarkers, which can be the first step in developing new therapies.

Perhaps the most promising area for ML in drug discovery is target/lead compound identification. The deep and narrowly focused data sets required for precision medicine and genomic-based therapies are almost ideally suited for ML-generated insight. The key is formulating the right questions. Typically, these questions need to be narrowly focused – such as: What are the biomarkers associated with the particular response?

By training algorithms on such data and formulating the right questions to ask of the data, firms can expect to realize more value from ML in drug discovery. It is important to remember, however, that relevant data in abundant quantities is a prerequisite for success. To help improve the possibilities for success in this regard, it is therefore important for the industry to work across the boundaries of individual firms and public agencies to make available as much data as possible.

Fortunately, there are promising signs of exactly this sort of cooperation and collaboration taking place. Organizations such as TransCelerate[6] emphasize intra-industry collaboration to speed drug discovery while organizations such the Accelerating Medicines Partnership (AMP)[7] and Accelerating Therapeutics for Opportunities in Medicine (ATOM)[8] both seek the same through public-private partnerships. Initiatives like these should receive the support they need to help drive drug discovery forward.

Portfolio optimization: Making investment decisions based on insight

R&D portfolio managers and those with roles responsible for managing the overall drug portfolio will want to evaluate how ML can optimize investment decisions. Beyond the realm of a single human's intuition and analytical capabilities, ML allows one to sort unwieldy amounts of data to reveal often surprising insights that can impact portfolio decision-making.

Only portions of the relevant data will be found in internal systems. Portfolio managers will also want to see government data on current clinical trials (in-house and with competitors), market data on demand for new therapies, and a wide range of other sources. Bringing all relevant data together in the cloud is the first step forward – followed by the application of machine learning to yield new insights.

It is not that human intuition and judgment are suddenly rendered useless; that's not the point. Instead, savvy managers increasingly regard ML as an essential tool for supplementing traditional research and analytical methods.

Ideally, ML ought to be seen as akin to obtaining a second opinion for making critical decisions on whether to accelerate or decelerate investments, or identifying gaps and opportunities in the product pipeline and portfolio. For example, ML can be used to identify milestones in one development project and evaluate them compared to other projects, internal or external. Pattern detection in the data can pinpoint trends based on a wide range of factors that could have greater predictive power than models currently used.

Clinical trials: Optimizing outcomes and improving success rates

It is said[9] that the success of Phase III FDA clinical trials is on the rise. One reason may be that firms are getting better at making go/kill decisions, thus foregoing Phase III trials all together for compounds that are unlikely to make it through to approval.

The ability to make these decisions is increasingly dependent on generating as much data as possible out of the Phase I trial. These data need to be of the proper shape to meet the needs of investigators. The requirement is narrow and deep data not wide and shallow. Moving forward, data meeting this requirement will enable firms to use ML to make better, more informed decisions.

In the face of growing volumes of data that are beyond the capacity of even the most expert teams to navigate, the case for ML to yield insights that would otherwise be overlooked is driving interest and adoption. In areas such as differentiating the characteristics of tumors and classifying them according to disease categories, ML is, in practical terms, the only available method for moving through high-volume data sets in a reasonable period of time. Indeed, as ML has emerged, it is increasingly regarded not as adding complexity to the clinical trial process, but rather as an invaluable tool to address the inherent complexity of the task at hand.

As always, investigators need to ask the right questions of the data, which in turn need to be of adequate quantity and quality. This challenge can be overcome with platforms that can connect to and can consume almost any kind and volume of structured and unstructured data in the cloud. The ease with which investigators can connect to diverse data sets will help to address the quantity issue, while a platform that facilitates a culture of data governance will help to ensure quality.

Patient profiling is among the most promising areas where ML can help in clinical trials. As precision medicine takes off, firms need trial participants with increasingly specific pathologies and biomarkers. With access to data sets from electronic medical records (EMR), investigators can now parse high volumes of patient data – creating synthetic cohorts, running simulations, and finding the exact matches they need to better ensure successful outcomes.

Because therapies are increasingly targeted to patients with specific genetic compositions, such as certain kinds of breast cancer where a patient has the HER-2 gene, firms must dramatically reduce the pool of trial participants to fit the prescribed genetic profile. This presents both a patient recruiting challenge and an opportunity to show statistical significance in the trial endpoint with a smaller enrollment that reduces cost and accelerates the time to approval.

One example[10] of ML for precision medicine comes from a group of researchers in Finland and Sweden. These researchers used ML to put more than 15,000 recently diagnosed diabetes patients into groups based on different disease characteristics and rates of progression. This sort of classification can make it easier to design therapies more likely to work for patients of particular circumstances.

The holy grail for ML in clinical trials may well be the notion of "virtual clinical trials," though some confusion has emerged as to the precise meaning of the term. Some refer to virtual clinical trials as the idea of "remote trials," where trial participants never actually come to a physical site but are monitored from afar using telemedicine, wearable devices, and sensors that transmit patient data. For precision medicine in particular, where it may be difficult to find patients with the specific set of biomarkers targeted for a therapy, virtual trials in this sense show great promise.

In another, perhaps even more promising sense, the term refers to the ability to simulate certain aspects of a trial in situ vs. in human. The idea of a "lab on a chip," for example, can support processes such as high-throughput screening using ML. The objective is to generate data that can demonstrate safety and efficacy for therapies at earlier stages of development, before clinical trials.

In still another sense, some refer to virtual clinical trials as the practice of running models with the aim of fully predicting outcomes. While few scientists in the life sciences see ML as replacing clinical trials, the technology is increasingly being seen as another tool to aid insight and decision-making. From this perspective, "predictive" virtual clinical trials can help investigators better plan trials and, as discussed above, identify smaller trial populations where a precise constellation of disease characteristics is targeted.

The life sciences supply chain: Controlling quality with predictive insight

Throughout the drug development process, from initial experimentation to animal and human trials, the inbound life-sciences supply chain remains a concern and a challenge. Ensuring quality for raw materials used in development and trials is non-negotiable.

As with research, the supply chain for life science companies generates a high volume of data, which can be analyzed to maintain the integrity of experiments, ensure safety for tests with animals and humans, and establish a sound repeatable process going forward into clinical trials.

ML can also be used to head off quality issues by revealing insights regarding which suppliers are most likely to deliver the right raw materials at the expected quality on time in given circumstances. This can make a dramatic difference. Even the smallest discrepancies in the quality of raw materials can put the health of patients and the success of a trial at risk.

Mishaps have already happened, such as substandard raw materials, inaccurate quantities of active ingredient, or failure to homogeneously disperse the active ingredient throughout the pills. These mishaps can come at the cost of billions of dollars.

One firm used raw materials that, when formulated, produced pink pills instead of white pills. On top of this, the active ingredient was incorrect (over 10X the concentration specified). It was also concentrated on one side of the pill, such that a patient on half a pill per day would receive no medication one day and too much the next. This firm had to shut down manufacturing, losing billions for what, in the end, amounted to a quality issue with inbound raw materials.

Purely as a risk mitigation tool, thus, ML is worth exploring when it comes to the life sciences supply chain. Using the technology to identify and select the best suppliers is an important component in ensuring patient safety and ensuring positive outcomes for the development and trial process.

None of this is to ignore more common application of ML for supply chain efficiency, including applications that are already in use in other industries. Firms, for example, can also use ML to predict disruptions in the supply chain based on diverse data, much of it publicly available. Data on weather patterns, regulatory changes, political and cultural events, or pollution levels in local streams – all of this can be incorporated into the mix to help drive better supply-related decision making. Firms adopting these approaches are exposing their stakeholders and trial participants to less risk.

Strategies for moving forward

While ML is undeniably a promising solution to the problem of data complexity in the life sciences, it is not a panacea. Some life science companies have invested tremendous resources in ML and AI in general, but have failed to realize the anticipated gains in R&D productivity because they underestimated the value of internal know-how and the need to integrate technical tools with decision making, core processes, and anticipated business results. Domain expertise is a prerequisite. Teams will continue to need to understand the industry, the processes for R&D and clinical trials, and the regulations that govern it all.

Do not trust machine learning in life sciences to technicians who lack foundational knowledge and industry experience. Subject matter experts need to work with technical teams to write the correct algorithm that asks the right question against high volumes of relevant data. After training the algorithm on relevant data sets, humans are needed to then look at the output and test hypotheses in an iterative fashion before the algorithm is finalized and put into production.

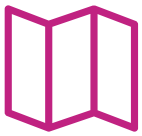
Strategies for moving forward

For firms to be successful, it is critically important to follow some key best practices designed to maximize the potential for success. Here are just a few:

Choose the right partner: Carefully research and vet the best partner for the overall project, from an industry as well as a technology perspective. Pick a partner that can work across fragmented landscapes and incorporate all data types, while also understanding the meaning of the data being analyzed. Also, be aware that the actual implementation of ML algorithms (as opposed to their design) is a common stumbling block. Look for a partner that knows how to put ML into action.

Build a strong internal team: An internal team that is dedicated to the project and has the executive backing to move forward is critical. The external team can only move as fast as the internal team. A dedicated internal team will also minimize time spent on irrelevant activities and help avoid common project pitfalls.

Set realistic, measurable goals: Throughout any ML project, teams need agreed-upon metrics and milestones to keep moving forward. Setting project goals based on these milestones helps to keep projects aligned and on track.



Develop roadmaps: Roadmaps for the initial project, as well as the longer-term journey, help to guide teams and the entire firm toward greater use and acceptance of ML. With a vision for what is to come, and with some idea of the advantages to be had, teams can work together in pursuit of the common objective.

Communicate: Throughout the project and the larger journey, it is best to err on the side of overcommunicating. Email and memos work, but even more effective is in-person communication. Encourage the sharing of information and explore alternative means of communication, such as videos or podcasts. This, too, will help teams stay aligned and on schedule.

References

[1] https://www.forbes.com/#12042f1c2254

[2] https://www.nytimes.com/2017/09/11/health/cancer-drug-costs.html

[3] https://www.sciencedirect.com/science/article/abs/pii/S0167629616000291

[4] https://www.phrma.org/Report/Biopharmaceutical-R-and-D-The-Process-Behind-New-Medicines

[5] https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-new-strategies-modernize-clinical-trials-advance

[6] https://transceleratebiopharmainc.com/

[7] https://www.nih.gov/research-training/accelerating-medicines-partnership-amp

[8] https://atomscience.org/

[9] https://blogs.sciencemag.org/pipeline/archives/2019/05/09/the-latest-on-drug-failure-and-approval-rates

[10] https://www.ncbi.nlm.nih.gov/pubmed/29503172

About SpringML

At SpringML, our domain experts in life sciences and machine learning enable Fortune 500 and emerging growth companies to extract value from their existing data. Our goal is to help customers move from data to knowledge to wisdom. In doing so, we partner with technology pioneers Google and Salesforce to help firms in the life sciences to develop analytical tools that access, process, and analyze internal and external data to improve discovery research, speed clinical trials, and make the supply chain for raw materials more reliable.

For more information

www.springml.com

1.800.346.8260