BCG

EARLY-STAGE INVESTMENTS COULD USE THE AI TOUCH

By Alexander Aboshiha, Michael S. Ringel, Wendi Backler, and Joseph Brilando

Early-stage investment opportunities are notoriously opaque. During the period when companies or inventors have started developing a technology, product, or service but are not yet ready to go to market, it's challenging to determine which innovations are most likely to succeed. This is particularly true in industries such as biopharma, where discovery and investigation of drug targets can start decades before success is evident in the form of financial returns. Investors use a combination of specific metrics and other, more qualitative information in their attempt to identify the companies that are likely to shine. However, the data are highly complex, and it's often unclear which metrics best correlate with success.

These challenging conditions create a fantastic opportunity. There is a wealth of leading-indicator metrics that we have strong reason to believe can help predict success. But exactly which ones matter most and how they interact with one another is difficult for human intelligence to decipher. As artificial intelligence (AI) tech-

nologies mature, a solution may be in sight. Our research shows that applying machine learning (ML) to prerevenue criteria can improve on the ability of traditional methods to determine the probability of a startup's success. In addition, it allows investors to speed up the process and screen a much larger number of potential opportunities.

The value of this approach is significant. In today's business environment, being able to make investment decisions with even just a little more certainty—whether that means seizing a good opportunity or avoiding a bad one—could be worth billions of dollars.

Identifying Early-Stage Opportunities by Conventional Means

Finding the best early-stage opportunities to invest in—whether startup companies, early drug targets, or emerging technologies—is a challenging proposition. Because each assessment requires a great deal of time, most teams can realistically assess only a small number of companies in any

given year. A large biopharma company, for example, may look at only 25 potential targets in a year, of which only one or two may turn out to be worth pursuing. Furthermore, the sources of these investment opportunities are often ad hoc, coming from investment bankers or company executives who happen to discover them via their networks or at conferences or trade shows.

In recent years, the haystack hiding the needle has grown significantly larger. The volume and complexity of the data that must be sifted have grown exponentially. As a result, business development (BD) teams can't maintain deep visibility into all the potential early-stage investments that are out there, and they struggle to make timely and informed investment decisions.

Building a Predictive Model
AI can help. In the traditional target-screening process, potential investors rely on the results of different kinds of analyses, along with their own business intuition. Complementing this human approach, ML algorithms can scan data to determine which variables contribute the
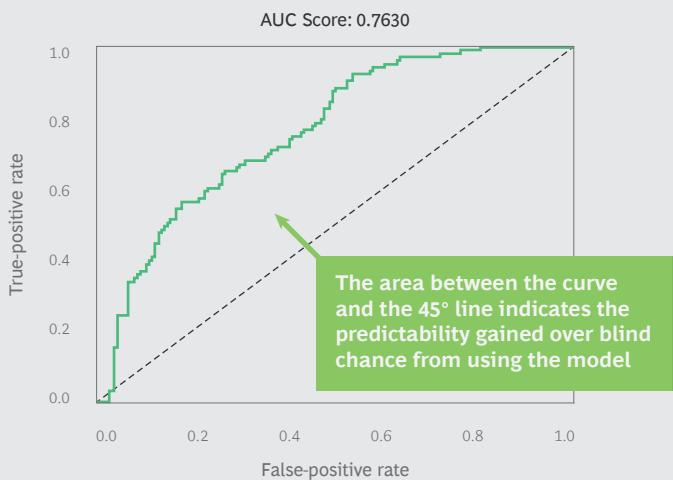
most to overall success, account for nonlinearities and the interactions between variables, and rank opportunities accordingly. As long as there is a sufficiently large training set, AI can elucidate nuances that elude humans.

The key lies in building a predictive model that is based on a robust train-test methodology. An example: we developed a model of success for nearly 300 immuno-oncology companies—using only data available in 2014—to predict how these companies would perform in 2017. When applied to a separate test set, the model successfully identified both over- and underperforming companies 70% of the time, which is significantly better than chance. Our model also had a very good "area under the curve" (AUC) value—a measure of predictive power—of 0.76. (See Exhibit 1.)

To be effective, such a model must be built on a relevant and meaningful definition of success. Given that some conventional financial metrics are irrelevant for prerevenue companies, we first trained and tested the model using total shareholder return (TSR) as the success variable and correlated current success with historical perfor-

EXHIBIT 1 | A Good Predictive Model Identifies Promising Investments Far Better Than Chance

Receiver operating characteristic (ROC) curve

AUC Score: 0.7630

The area between the curve and the 45° line indicates the predictability gained over blind chance from using the model

0.76

Area under curve (AUC)
Quantification of a model's predictive power over blind chance—an AUC score of 0.5 indicates that the model has no predictive power

Sources: BCG Center for Innovation Analytics; BCG analysis.
Note: An ROC curve compares two operating characteristics (true-positive rate and false-positive rate) as the criterion changes. The curve shown is for BCG's model of success for immuno-oncology companies. The AUC represents the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

mance on prerevenue metrics. This makes it possible to apply the model to companies today on the basis of these same metrics and assess their likelihood of success in the future, even if the companies are not yet generating any revenue.

Which measures most strongly correlate with future success? Interestingly, although we initially included both financial and nonfinancial metrics in our model, we found that nonfinancial metrics are particularly helpful for uncovering previously unknown opportunities. These metrics include data regarding publications, grants, patents, and similar variables. Of these prerevenue metrics, indicators of scientific acumen are the strongest contributors, followed by measures of influence and technology. (See Exhibit 2.)

Many of these metrics are the same kinds of factors that people weigh when assessing a company or other investment opportunity. The difference is that people do it in an intuitive way based on a limited set of prior comparables, whereas AI employs a quantitative algorithm based on a large training set. This doesn't mean that BD departments should jettison their experienced staff in favor of AI. Nor is it fair to say the opposite. As in many other do-

mains, the best results to date have been achieved by combining the inputs from both sources.

## Casting a Wider Net at an Earlier Stage

The advantages of combining artificial intelligence and prerevenue metrics are considerable. This approach not only enables companies to screen a larger number of early-stage assets much faster and more cheaply than in the past but also lets them use a much wider range of criteria in the screening process.

This method is applicable to many different use cases, all of them with the potential to create tremendous value, especially in technology-driven industries:

- Monitoring progress in early stages of product development

- R&D prioritization and allocation of resources

- Sorting false positives from likely successes (in assessment of potential products)

- Early-stage licensing and investment

EXHIBIT 2 | The Predictive Power of Prerevenue Metrics

Overall, 80% of predictive power comes from 25 variables…
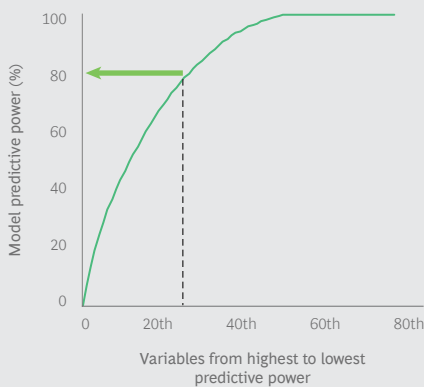
…21 of which point to the predictive power of scientific acumen and influence
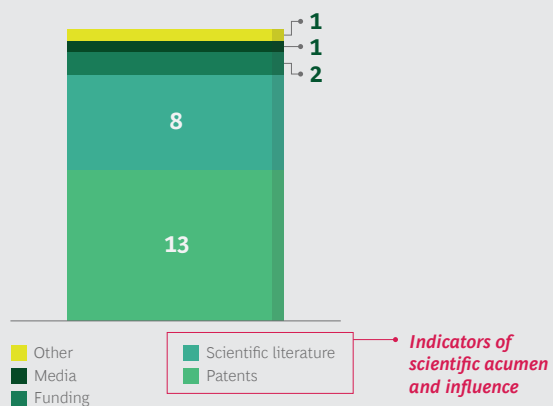
Sources: BCG Center for Innovation Analytics; BCG analysis.
Note: Data from BCG's model of success for immuno-oncology companies.

- Creation of options, either progressive or full acquisitions

- Valuing late-stage and in-market companies

- Identifying investment opportunities

A biopharma company, for example, applied the technique to predict which combinations of biological targets and clinical indications would most likely lead to successful drugs in cancer. The predictive model, which was trained on about a million data points from hundreds of target-indication pairs for which the outcome of drug progression was known, demonstrated again that variables related to scientific acumen have the greatest predictive power.

Adding AI to the BD Toolbox

There's no better time to get on the AI bandwagon. Thanks to greater availability of and access to data, cloud-based storage options that reduce the need for internal infrastructure, and statistical packages that are in easy-to-use languages, companies can deploy ML to their advantage.

But before getting started, it's critical for companies to determine whether they have the capabilities to integrate AI into their business development and target identification processes. Three prerequisites are especially important.

Alignment with Overall Business and Digital Strategy. Before launching any AI initiative, companies first need to be clear on their overall goals and expectations for using it, and how it fits into their overall business and digital strategy. It's essential to position AI among the other tools in the digital toolbox. AI complements and supplements the deep topic expertise and relationships of the BD and R&D teams—it does not replace them.

Access to the Right Data and Digital Tools. Companies need to make sure that they have the tools and infrastructure required for working with the most up-to-date data, as a recent BCG article explains. Many

companies have relatively little process digitization and lack a systematic approach for leveraging partner ecosystems and their wealth of data.

Equally important, companies must be able to incorporate the results of their digital analyses into a robust process for screening opportunities. They also need to be able to incorporate these results into the internal processes used to make investment decisions. This requires engagement with, and buy-in from, experts and seasoned decision makers who are willing to learn about the approach, apply it appropriately, and trust the insights.

The Right Kinds of Expertise. Data scientists are required, and it's important that some of them have years of experience working with the pertinent types of data and analytics.

But data scientists alone are not enough. Also important are subject matter experts and experts in innovation analytics—people who understand the nuances involved in coming up with meaningful prerevenue success indicators, who are adept at finding ways to measure some of the less quantifiable factors, and who can sense-check and help adjust models over time.

In addition, it's critical to have people who can interpret the results from a strategic perspective. For example, understanding the variables that contribute most to the predictive power of the model can help define action steps for ongoing monitoring of new developments. In many fields, such as baseball prospect selection and weather forecasting, the best results are often obtained using a combination of human and machine intelligence. We believe the same holds true in early-stage investing.

D ESPITE THE PROMISE of machine learning, it has yet to gain wide traction in business development circles, though dozens of venture capital and private equity firms are using technology for investment decisions, and many startups are going to market with ready-built models. But there

are only a few companies currently deploying machine learning for business development purposes.

Our research clearly demonstrates that using machine learning to analyze nontraditional metrics, in conjunction with traditional scouting and target identification activities, can greatly facilitate the prioritization of early-stage opportunities that hold the highest likelihood of success. The process is challenging, requiring much time and many iterations, and there is no guarantee of acquisition success. But all told, adding machine learning to the mix is likely to make it easier, faster, and cheaper to find the needle in the ever-growing haystack. Even if it improves speed and decision making by only a small fraction, that's enough to be hugely valuable in an environment where companies are spending hundreds of millions of dollars to get a competitive edge.

About the Authors

Alexander Aboshiha is a managing director and partner in the Los Angeles office of Boston Consulting Group and works across a wide range of practice areas, focusing primarily on health care and digital health, corporate development, and organizational transformation. He is a member of BCG's global digital health leadership team. You may contact him by email at aboshiha.alex@bcg.com.

Michael S. Ringel is a managing director and senior partner in BCG's Boston office and is the global leader of the firm's Center for Innovation Analytics. He works extensively in health care, covering a range of strategic, digital, and operational topics. You may contact him by email at ringel.michael@bcg.com.

Wendi Backler is a partner and associate director in the firm's Toronto office. She specializes in intellectual property, open innovation, and innovation networks. She founded BCG's Center for Innovation Analytics, which uses data science and big data to unearth emerging trends. You may contact her by email at backler.wendi@bcg.com.

Joseph Brilando is a senior advisor to BCG. As a venture capitalist in the 1990s, he supported the B2B internet revolution. He is based in the firm's South San Francisco office and the founder of its ValueScience Center. You may contact him by email at brilando.joseph@bcg.com.

Acknowledgments

The authors thank Nicole Quenneville, Harsh Kaushik, Ruchika Mendiratta, and Brett Schiedermayer for their analytical support and Julian King for his expert guidance.