Harnessing Machine Learning for Due Diligence: Realizing the Possibilities

A wave of technology solutions driven by advances in artificial intelligence promises to revolutionize due diligence. However, it's essential to keep expectations realistic and to know how your machine learning program learns.

The increased emphasis on due diligence and the ever-growing amount of open source and proprietary data available on due diligence subjects combine to create an ideal use case for machine learning technology. While automated due diligence platforms offer tantalizing possibilities, they can also lead to frustration and unfulfilled expectations. Organizations considering these solutions can greatly increase the chances of success by approaching implementation holistically and by knowing how to evaluate technologies critically.

Harnessing Machine Learning for Due Diligence

WHAT TECHNOLOGY CAN—AND CANNOT—DO

As with all technology, implementation of a due diligence platform powered by machine learning needs to begin not with the technology but with the larger context of improving the function itself. This means starting with a comprehensive review of the due diligence workflow. What are the regulatory or best-practice requirements that must be met? How are data and risk assessments about customers and other third parties shared across the organization? How adequate is the response mechanism to identified risks? Mapping the overall due diligence function and identifying gaps and bottlenecks will provide a blueprint for progress. Some of those gains will be powered by technology, but others will require changes in processes or capabilities. For example, a due diligence platform may help an institution increase the throughput volume and the consistency of risk ratings, but achieving meaningful gains in due diligence effectiveness may also require thorough data remediation and a clarified risk escalation framework. Making technology part of a larger solution thus allows the institution to specify its technology requirements—and expectations—with greater precision. That solution also should reflect the institution's overall preference for either building in-house compliance capabilities or outsourcing them.

After determining the requirements for the technology, the enterprise must factor in perspectives from its various divisions. The IT department's view will be based on how the due diligence technology needs to integrate with existing systems. The cybersecurity team will need to ensure that no vulnerabilities are being introduced, and the finance department will want to know the expected return on investment.

DARREN BURRELL

Vice President
Compliance Risk and Diligence
Reston, VA, US
darren.burrell@kroll.com

PEERING INSIDE THE BLACK BOX

These steps provide a framework for establishing the platform's functional requirements, but that is only part of the equation. Organizations must also be able to evaluate the technology itself, a task made all the more challenging by the ubiquity of the term *machine learning* and the absence of a clarifying legal standard for it. Consequently, organizations evaluating due diligence platforms need to be sure they understand exactly what those products deliver. Such understanding is critical because an application's inner workings directly determine the volume, accuracy and speed it will achieve under real-world conditions.

The most common form of machine learning uses what is known as *supervised learning*. In supervised learning, an algorithmic model is fed large amounts of historical data and seeks predictive patterns. For example, it might use data on the size, location, amenities, and sales price of homes. As the model analyzes the data, it attempts to predict the sales price of each home, checking its prediction against the actual price information that is included in the dataset. With each prediction it makes, the model fine-tunes itself until it can satisfactorily predict a home's sales price based on the other variables. In the due diligence context, a model might be used to identify and classify risk-relevant information, reduce false positives when researching against open source data or assign a money laundering risk score to a customer based on transaction history, currency used, industry, jurisdiction and other attributes.

Two key takeaways emerge from this overview. First, while human programmers necessarily revise the learning algorithm to improve its accuracy, the prediction process itself occurs with no outside intervention. This defining characteristic of true machine learning is an essential criterion in any product evaluation. Some due diligence programs that claim to be driven by machine learning actually use low-cost labor through platforms like Amazon Mechanical Turk to make predictions by applying simple checklists.

Second, the quality of the algorithmic model is largely determined by the quality and quantity of the data used to train it. Indeed, this explains why firms like Google and Facebook distribute machine learning programs as open source software: These companies use the massive amounts of data their programs collect to refine the proprietary machine learning models they use internally. The data is actually more valuable than the algorithms themselves, because of its volume and because, being naturally generated, it reflects the nuance and randomness of the real world. Thus, for due diligence models, training datasets collected by analysts in the course of research and discovery are superior to datasets that have been artificially assembled. Naturally-generated datasets represent real-world scenarios more accurately while also capturing the thought processes of the expert analysts who compiled the data during their due diligence work.

Machine learning technology can be a powerful component of an organization's due diligence arsenal. However, enterprises considering using such a tool need to specify its role in detail and to subject its internal workings to careful review.