

Reality Check: The Use of Big Data and Predictive Data Models

日本語版も入手可能
www.isaca.org/currentissue

Humans have long been enamored with the idea that if they can just feed enough data about the past into a machine—whatever that machine is—it can predict what will happen in the future. In a 1984 episode of the US animated television show *The Transformers*, the Autobots look for an elusive space bridge to their home planet, Cybertron. Their human friend, an archetypal whiz-kid named Chip Chase, informs them that:

[B]y feeding Teletraan 1 [the Autobots' supercomputer] all the data we have about the space bridge's last appearance, I might get it to predict where the bridge will appear next.¹

It works. If only reality were as straightforward as cartoons.

The idea of using vast amounts of data to anticipate what is going to happen is not new. However, the power, precision and affordability of predictive models are increasing based on the increasing volume and availability of data in general and the computing power of the cloud. More organizations are going to be faced with tough questions and issues related to the real-world use of such models—discussions that may previously have taken place on a more hypothetical level.

For many enterprises...investing in predictive data modeling is a worthwhile pursuit that can increase customer satisfaction, improve efficiency and even help save lives.

The notion that every organization is sitting on a veritable crystal ball in the form of untapped data is a fantasy. However, for many enterprises across a broad range of industries, investing in predictive data modeling is a worthwhile pursuit that can increase customer satisfaction, improve efficiency and even help save lives. But to realize this value, organizations must deal with some down-to-earth and sometimes messy realities.

Data Security and Integrity

Cybersecurity remains at or near the top of every organization's risk register. The threats to data security and integrity are increasing from within organizations (e.g., complexity of data governance, conflicting priorities, neglect) and from the outside (e.g., malicious acts, lack of control over third parties). At the same time, the stakes accompanying a data breach are growing in terms of potential financial and reputational damage and legal/regulatory liability.

KEVIN M. ALVERO | CISA, CDPSE, CFE

Is senior vice president of internal audit, compliance and governance at Nielsen Company. He leads the internal quality audit program and industry compliance initiatives, spanning the enterprise's global media products and services.



According to a 2022 Protiviti report:

IT audit teams, as well as other departments (e.g., legal, compliance, IT), are scrambling to keep pace with new data privacy and data security rules as well as changing legal and regulatory compliance requirements that have growing implications for organizational data management and technology-related activities.²

Particularly as it relates to sensitive data, the fundamental question is whether the value the organization is getting from these data is worth the risk of ownership. To answer that question, top leadership must have a clear and shared understanding of how predictive modeling is expected to support the organization's mission, strategy and core values. At the same time, the capacity for predictive modeling to impart value to the organization is directly related to the quality and integrity of the data that are input into the models. Therefore, a strong sense of purpose and a commitment to data security and integrity must be in place from the top down for organizations to avoid dabbling in, or lunging after, the prospective benefits of predictive data modeling in a manner that puts the organization at excessive risk.

Bias, Privacy and Other Ethical Concerns

Members of the general public have become increasingly concerned that the personal data organizations collect (with or without consent) will be used in ways that violate their right to privacy or their right to fair and equitable treatment. Anticipating people's thoughts and actions too well can be downright creepy, and it can negatively impact their perception of a brand and its level of trust. Voicing that concern has led to change, both in government and in the marketplace. A 2022 *Harvard Business Review* article notes that:

Until now, companies have been gathering as much data as possible ... often without customers understanding what is happening. But with the shift towards customer control, data collected with meaningful consent will soon be the most valuable data of all, because that's the only data companies will be permitted to act upon.³

In addition to the increased focus on privacy, there is a greater demand for transparency to ensure that

Leadership must be able to determine if available data are relevant to the prediction they are trying to make, and if it is worth the risk and cost of acquiring, accessing, storing and including that data in the modeling.

organizations that utilize advanced data analytics are treating people fairly and equitably. In particular, this concern is relevant to a type of predictive data modeling called clustering, in which data (and the people that data represent) are placed into various groups based on their common attributes. In addition to targeted advertising:

...other use cases of this predictive modeling technique might include grouping loan applicants into 'smart buckets' based on loan attributes, identifying areas in a city with a high volume of crime, and benchmarking [Software-as-a-Service] SaaS customer data into groups to identify global patterns of use.⁴

At one level, this seems intuitive and reasonable. If the purpose of advertising is to inform people about products and services they might want, then using information about those people to improve the odds of suggesting relevant products to them only makes sense. The same could be said about applying the known, historical likelihood of a destructive event occurring to the decision of how much a customer should have to pay for insurance against that event.

However, when organizations make determinations (even well-supported ones) about whether certain clusters of people can, should or would want to do certain things, they risk crossing the line between prudent risk management or beneficial tailoring of the customer experience and discrimination. The data economy, "was structured around a 'digital curtain' designed to obscure the industry's practices from lawmakers and the public ... [but] that curtain has since been lifted."⁵ Therefore, organizations must understand how their model-powered decision-making processes will tolerate this sunlight and, more important, consider proactively how their use of predictive data modeling aligns with their core values.



LOOKING FOR MORE?

- Read *Data Science Fundamentals Study Guide*. www.isaca.org/credentialing/itca/data-science-fundamentals-certificate
- Learn more about, discuss and collaborate on information and cybersecurity in ISACA's Online Forums. <https://engage.isaca.org/onlineforums>

Data Relevance

Forecasting is another common use case for predictive modeling. Although organizations have traditionally used historical data to anticipate demand, the comprehensiveness and immediacy of information that can be included in the calculation necessitates a new level of scrutiny over the relevance of data. In the past, an enterprise had, perhaps, transaction volume, returns data, customer loyalty program data, foot traffic statistics and some demographic data, but now there exists the ability to incorporate a much wider variety of variables into the computation with real-time trending. The fundamental problem thus shifts from needing more data to determining where to draw the line. Leadership must be able to determine if available data are relevant to the prediction they are trying to make, and if it is worth the risk and cost of acquiring, accessing, storing and including that data in the modeling on the premise that doing so could make forecasts even incrementally more accurate.

Organizational leaders must seriously consider whether they have, or can acquire, the skilled workers needed to execute their strategies for predictive modeling.

Disruption

It is also important for business leaders to fully comprehend the ongoing commitment to monitoring that is required for the responsible and effective use of predictive models. Such models cannot simply be deployed and left to run at the risk of developing biases that could impair their decision-making and expose the business to risk.

Ensuring that changes in the marketplace or the broader world do not break data models necessitates vigilance, but it also requires that resilience be an upfront consideration in the early stages of planning the development and use of a predictive model. In short, the model must be built with the assumption that the environment in which it ultimately performs will change frequently postdeployment in ways that may not have been foreseen during development. These are key factors in the return

on investment (ROI) equation that business leaders must understand before embarking on a predictive modeling initiative.

ROI

For organizations in certain industries, investment in advanced data modeling algorithms will almost certainly be worthwhile. However, the organization must still be able to quantify its ROI. The ability to better anticipate ROI can have numerous quantifiable benefits, including:

- Less waste resulting in lower costs
- Fewer inefficiencies and useless procedures
- Reduction in costly or dangerous delays
- Improved quality with fewer errors

For example, when it comes to emergency response or disaster preparedness, better predictive models can literally translate to lives saved. However, in other situations, the cost-value picture is less clear. For example, a taxi or ride-sharing service could benefit greatly from the ability to analyze vast amounts of consumer and marketplace data to predict—as accurately as possible—the timing of an event that will draw prospective customers and precisely how many. On the other hand, a brick-and-mortar storefront managing inventory levels can probably predict demand well enough simply by tracking historical sales data, meaning that investment in advanced predictive models is likely not worth the cost.

Moreover, customer demand for forward-thinking experiences that help them make better decisions and actions “doesn’t always suggest a predictive solution.”⁶

This speaks once again to the need for leadership to understand how increased accuracy and insight from predictive models can benefit the organization and, even better, to be able to quantify those benefits before investing in predictive technologies.

Organizational Skills and Expertise

It is expected that 97 million jobs involving artificial intelligence (AI) will be created between 2022 and 2025.⁷ “AI has the potential to transform every industry...however, businesses are still struggling to find employees with the skills necessary to create, train and work alongside intelligent machines.”⁸

Although demand for skilled workers in the field reportedly has been quickly outstripping supply, the problem of organizational skills and expertise related to predictive modeling is not strictly a hiring problem. According to research by the Massachusetts Institute of Technology (MIT) (Cambridge, Massachusetts, USA) Center for Information Systems Research:

Creating successful artificial intelligence programs doesn't end with building the right AI system. These programs also need to be integrated into an organization, and stakeholders—particularly employees and customers—need to trust that the AI program is accurate and trustworthy.⁹

This, the researchers conclude, is the case for building enterprisewide AI explainability.¹⁰

Organizational leaders must seriously consider whether they have, or can acquire, the skilled workers needed to execute their strategies for predictive modeling. They must also determine if they can raise the level of literacy within their organization to realize the full potential benefits of predictive modeling.

Conclusion

Anybody who has ever researched a potential investment opportunity has been somberly reminded that past performance is not indicative of future results. Nevertheless, for many organizations, gaining the power to anticipate the needs of individual customers ever more accurately and foresee shifts in the broader marketplace is worth tackling the unknowns and potential pitfalls associated with advanced predictive data models.

Research shows that:

Most large firms already suffer from a series of internal tensions over customer data...and up to 90 percent of current IT budgets are spent simply trying to manage internal complexities, with precious little money actually spent on data innovation that improves either productivity or the customer experience.¹¹

That type of dysfunction cannot be overcome by an algorithm, no matter how sophisticated. Rather, it requires commitment, a clear strategy and well-aligned goals to ensure that any predictive model is built on a solid foundation.

Endnotes

- 1 *The Transformers*, Season 1, Episode 6, "Divide and Conquer," directed by John Walker, written by Donald F. Cohort, 20 October 1984, syndicated
- 2 ISACA® and Protiviti, *IT Audit Perspectives on Today's Top Technology Risks*, USA, 2022, www.isaca.org/it-audit-2022
- 3 Rahnama, H.; A. Pentland; "The New Rules of Data Privacy," *Harvard Business Review*, 25 February 2022, <https://hbr.org/2022/02/the-new-rules-of-data-privacy>
- 4 Insightsoftware, "Top Five Predictive Analytics Models and Algorithms," 1 January 2022, <https://insightsoftware.com/blog/top-5-predictive-analytics-models-and-algorithms/>
- 5 *Op cit* Rahnama and Pentland
- 6 Blanchard, B., et al.; "Predictive Modeling and Influencing Customer Behavior," Microsoft, <https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/innovate/considerations/predict>
- 7 Marr, B.; "What Are the Most In-Demand AI Skills?" *Forbes*, 13 June 2022, <https://www.forbes.com/sites/bernardmarr/2022/06/13/what-are-the-most-in-demand-ai-skills/?sh=4682e7b3249c>
- 8 *Ibid.*
- 9 Brown, S.; "Why Companies Need Artificial Intelligence Explainability," Massachusetts Institute of Technology (MIT), Swan School of Management, Cambridge, Massachusetts, USA, 21 September 2022, <https://mitsloan.mit.edu/ideas-made-to-matter/why-companies-need-artificial-intelligence-explainability>
- 10 *Ibid.*
- 11 *Op cit* Rahnama and Pentland