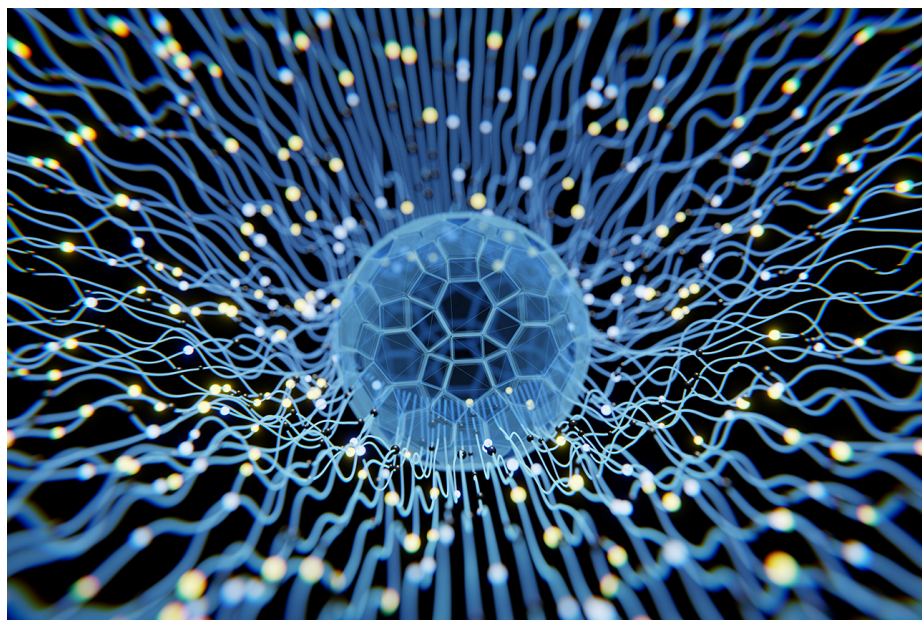# MITSloan
Management Review

Data Science, Quarantined

Companies are beginning to reboot their machine learning and analytics, which have been disrupted by the global pandemic.

Jeffrey D. Camm and Thomas H. Davenport • July 15, 2020

READING TIME: 7 MIN

The economic impact of COVID-19 is unprecedented, dramatically changing markets and prospects for economic growth. Supply chains, transportation, food processing, retail, e-commerce, and many other industries have transformed overnight. Unemployment in the U.S. has reached levels unknown in recent

memory, and GDP is expected to fall around the world. As one economic journalist summed up the situation: "Nearly everything in the world is super-weird and disrupted right now."

The data we use to make good managerial decisions has been caught up and turned upside down in this unpredictable marketplace. This is no small matter: Over the past decade, we have seen a dramatic movement toward data-driven decision-making, in step with an explosion of available data sources. Point-of-sale data, the internet of things, cellphone data, text data from social networks, voice, and video — are all automatically collected and reported. Coupled with advances in machine learning and artificial intelligence, these resources enable leaders and organizations to use analytics and data science for better-informed and improved decisions.

But what we're now evaluating is what happens to this accelerated, data-driven approach when a large-scale disruption, such as a global pandemic, results in a seismic shift in data. Machine learning models make predictions based on past data, but there is no recent past like today's present.

To better understand the impact on data science of our current moment and how the disruption will be managed going forward, we reached out to a number of data science and analytics directors. We asked what they have experienced in recent months and how they plan to adjust and redeploy their machine learning

models as organizations enter a new economic environment.

A Pivot to Fast-Cycle Descriptive Analytics

Every analytics manager we spoke with described the same basic reaction as the pandemic began to disrupt their operations: Regardless of whether the pandemic caused the demand for their company's products and services to plummet (as it did for, say, apparel) or to spike dramatically (for instance, toilet paper), there was an almost instantaneous shift away from more advanced analytics focused on prediction and optimization to descriptive analytics such as reports and data visualization. Descriptive analytics helped companies get a better understanding of what was happening.

Because of the volatility of the situation, all cycle times for reporting were dramatically compressed. The demand for real-time dashboards increased. As one manager from a global consumer goods company described it, "We weren't worried about detailed forecasting, we were just trying to get the shapes of the distributions right."

Dan Rogers, director of data science and operations research at 84.51°, a marketing analytics company owned by supermarket giant Kroger, echoed that.

"There were definitely a lot of resources applied to descriptive reporting at first as we strove to understand what was happening and how the pandemic was affecting our company," he said. "Entire teams were put on this, doing much of the same analysis they always did, but at an accelerated rate. A monthly or quarterly report might now be requested weekly or even daily." His teams have also done some descriptive modeling to help isolate the impact of the pandemic, he said. "This work can turn into predictive modeling to forecast the ongoing impact and better understand the 'new normal' we find ourselves in."

At some companies, data teams were asked to focus on specific pain points. At automaker Ford, executives have been less interested in commonly gathered report and dashboard analytics during the pandemic, said Craig Brabec, the company's director of global data insights and analytics. Instead, they are more likely to ask for custom analyses involving particular situations (for example, the extent of rail delays in the Mexican port of Veracruz) and new data sources.

Predictive Analytics and Automated Machine Learning Get Sidelined

Even in normal times, demand forecasting is one of the most difficult challenges for data scientists. Changing

consumer demand, volatile market conditions, and competitive moves all make predicting demand a trial. As the pandemic hit, structural shifts in demand wreaked havoc on machine learning models that were slow to adapt to the unusual data. As one manager quipped, "Our demand-forecasting automated machine learning models didn't handle eight weeks of zeros very well."

As companies shifted focus to descriptive analytics to understand changes in trends, they put their machine learning models for forecasting demand on hold. They started relying on simple forecasting approaches such as asking, "What did we ship yesterday?" or using time-series smoothing models such as computing moving averages, while closely monitoring the demand data to see if new patterns were emerging.

In the case of automated machine learning, many companies let their models continue to run, using the pandemic as a unique learning opportunity. By closely monitoring how the models were adapting to the unusual data, data scientists could better understand the robustness (or lack thereof) of the models. Lydia Hassell of apparel manufacturer Hanesbrands oversees over 100,000 machine learning models for product demand forecasting, and she says she utilized more frequent runs of machine learning exception reports. "These exception reports provide details on outliers from the machine learning models," she explained. "While we would normally run these reports on a

monthly basis, we began running these weekly, or even more frequently, to better monitor what was happening to the machine learning models." Hassel immediately started to use the reports to update and test new models to forecast into 2021.

Some companies attempted to use new external data sources to try to predict demand. Brabec at Ford said that in order to understand and predict consumer demand, analysts began employing aggregate connected-vehicle trip data that indicates either increases or decreases in driving activity nationally, as well as air pollution levels and car-related internet searches. "Some of this data may not be leading indicators of car sales," he said, "but they seem to at least move in parallel with them, and they suggest that the marketplace is opening up."

Other companies, lacking valid data for their models, simply made policies more conservative. This has been particularly true in credit risk models. Several banks, for example, raised the credit score requirements for home mortgages by substantial amounts: JPMorgan Chase, for example, raised the required credit score for new and refinanced mortgages to 700, and the minimum down payment to 20%. One analytics executive told us with regard to his company's credit models, "Those with 800-and-over scores are fine; everyone else is suffering. We model our customers as we did pre-COVID and add an extra risk factor."

Next Steps for Rebooting

How do we move forward with predictive analytics and machine learning given the disruption to data we have seen? What will the new data normal be, and how long will it take us to get there? Based on our conversations with directors of data science and analytics, we propose that the following should be considered as a part of a strategy in the near and long-term future:

Weigh data relevancy — what to delete, what to keep, what to impute. Should unusual data during the pandemic be deleted? Should it be replaced with imputed values based on data prior to COVID-19? Is pre-COVID-19 data even relevant going forward? The answer to these questions will surely differ by sector. Using moving averages — where you compute the average of a subset of data to balance out random fluctuations — and other smoothing forecasting techniques were mentioned by a number of the analytics managers as a way to navigate how much to rely on pre- and post-pandemic data.

Embrace increased use of external data. Trying to model low-probability, highly disruptive events will require an increase in the amount of external data used to better account for how the world is changing. The right external data could provide an earlier warning signal than what can be provided by internal data. One director whose company relied on the Johns Hopkins University COVID-19 site for data stated that a new

metric of effectiveness could be to consider how fast external data can be integrated into existing systems for use in analytical models.

Ramp up model auditing and stress testing. Several analytics leaders mentioned the need to keep close tabs on their machine learning and prescriptive models. They said they planned to audit data input, model assumptions, and model output more frequently. How will models respond to zero demand, tenfold demand, or anomalies like the negative price of oil? Techniques developed for quality control in industrial engineering, like control limits and acceptance sampling, need to be applied to machine learning to make sure the models are "in control."

Construct a portfolio of specialized models. A consumer goods manager mentioned that once his company's data team better understood what was happening with the pandemic, they started deploying their hurricane models. Consider developing scenario planning and simulations to construct specialized models that can be "pulled off the shelf" as needed. What have you learned since the outbreak of COVID-19 that you could implement if there is a second and perhaps worse wave of infection next winter?

Everyone we spoke to mentioned the shortened cycle times for model development and deployment. One person said the new normal for data science will be "all about agility and speed." The ability to generate customized and adaptive models quickly will be a key

determinant of success: It's a different world from the relatively stable data and analytics world of the past. As one director of analytics commented, "We'd better get used to operating with bad data for a while."

Topics

| Data & Analytics | | Leadership | | Analytics & Performance |

| IT Governance & Leadership |

ABOUT THE AUTHORS

Jeffrey D. Camm is the Inmar Presidential Chair in Business Analytics, associate dean of business analytics, and executive director of the Center for Analytics Impact at the Wake Forest University School of Business (@wakeforestbiz). Thomas H. Davenport (@tdav) is the President's Distinguished Professor of Information Technology and Management at Babson College, as well as a fellow at the MIT Initiative on the Digital Economy and a senior adviser to Deloitte's Analytics & Cognitive practice. Their recent article, "The Recession's Impact on Analytics and Data Science," coauthored with Melissa R. Bowers, explores the demand for analytics and data science resources.

TAGS: Analytics Strategy, COVID-19 Resources, Data Science,

Digital Leadership, Machine Learning