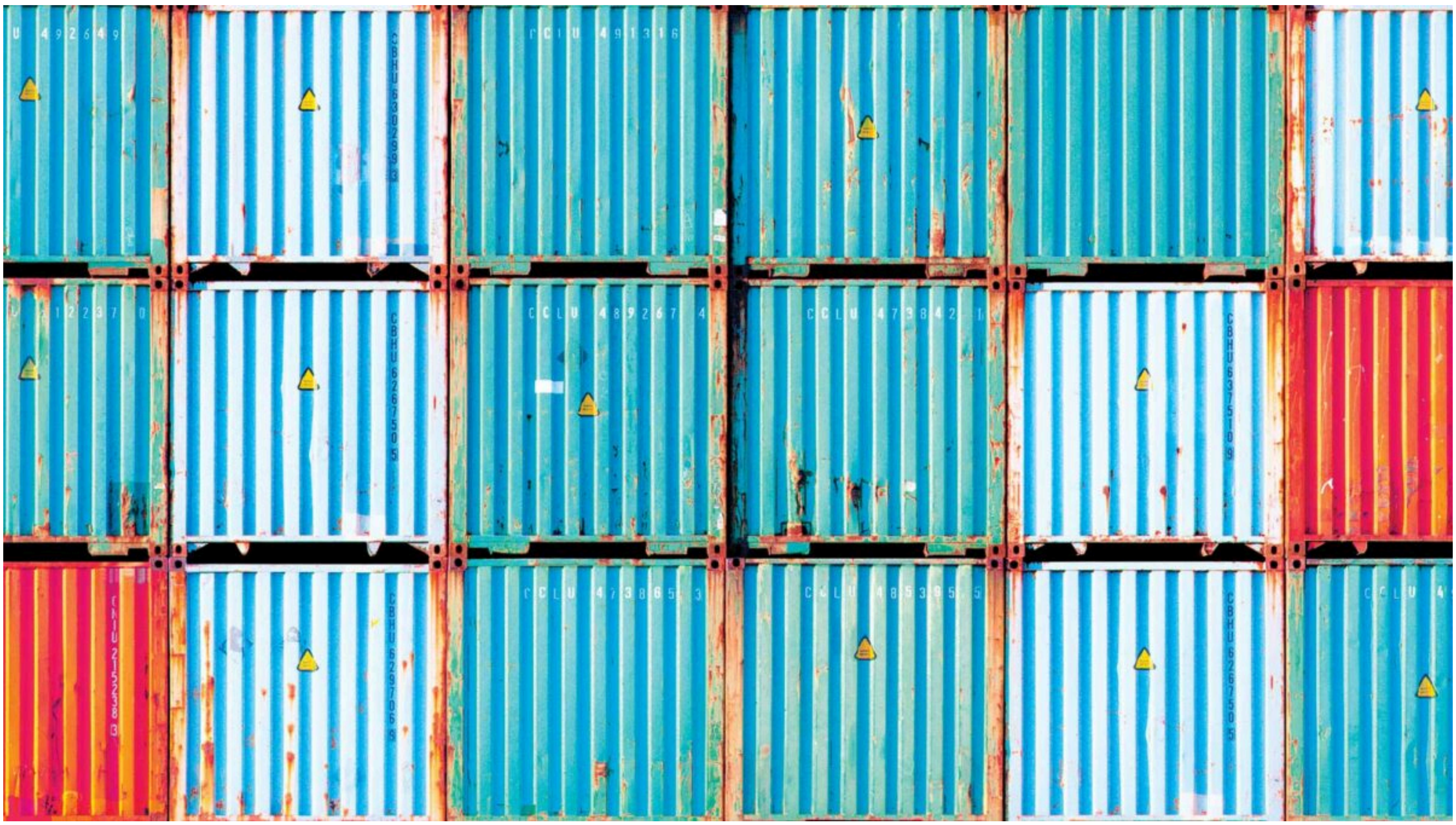
Harvard Business Review

Supply Chain Management

How Machine Learning Will Transform Supply Chain Management

by Narendra Agrawal, Morris A. Cohen, Rohan Deshpande, and Vinayak Deshpande

From the Magazine (March–April 2024)



Didier Engels

Summary. Businesses need better planning to make their supply chains more agile and resilient. After explaining the shortcomings of traditional planning systems, the authors describe their new approach, optimal machine learning (OML), which has proved effective in a... **more**

The Covid-19 pandemic, the Russia-Ukraine conflict, trade wars, and other events in recent years have disrupted supply chains and highlighted the critical need for businesses to improve planning in order to be more agile and resilient. Yet companies struggle with this challenge. One major cause is flawed forecasting, which results in delivery delays, inventory levels that are woefully out of sync with demand, and disappointing financial performance. Those consequences are hardly surprising. After all, how can inventory and production decisions be made effectively when demand forecasts are widely off?

We have developed a way to address this deficiency. Our new paradigm uses machine learning and historical data to generate superior recommendations for supply chain decisions. While current machine-learning methods focus on trying to create more-accurate forecasts, ours focuses on making actual decisions. This new methodology, which we call *optimal machine learning* (OML), involves using artificial intelligence technology to create a mathematical model that takes key data inputs related to the supply chain (the nodes of the network, their locations, sales and shipment transactions, financial parameters, marketing promotions, logistical and capacity constraints, and so on) and links them to planning decisions (what quantities to produce, for example, or what levels of inventory to stock at each location). This model can take into account a company's priorities (such as the level of customer service it is contractually obligated or wishes to achieve), its budget restrictions, and other resource constraints (such as the availability of materials and labor). The data is stored in a way that enables updating in near real time and quick revision of the calculations that inform decision-making.

We developed OML after decades of researching supply chain management and implementing our resulting ideas at companies in a range of industries: semiconductor equipment manufacturing, aerospace and defense, telecommunications, and computing. In this article we explain why existing supply-chain-planning methods, including other machine-learning techniques, have failed, how our approach works, and what changes in planning teams and metrics are needed to implement it. We highlight our experiences putting OML to the test at two *Fortune* 500 companies where we served as consultants—and where the results demonstrate the potential to increase revenue and product availability with significantly lower investments in inventory.

One of the companies makes and sells capital-intensive equipment for manufacturing semiconductors. To ensure that its customers can keep the equipment up and running, the firm must manage an inventory of thousands of spare parts. In the past, though, it often faced shortages of key components, which led to expensive expedited shipments or disruptive wait times that caused costly outages at customer locations. Frustrated managers frequently scrambled to override the inventory recommendations offered by their enterprise resource planning (ERP) system. The managers relied on their own experience and used data inputs not explicitly incorporated in the existing planning software, but their overrides were subjective, ad hoc, and time-consuming to devise, often compounding their problems.

The other company we'll discuss is a consumer electronics firm that sells, through thousands of retailers' stores, a portfolio of advanced virtual-reality interface devices produced by contract manufacturers in Asia. This firm was building up its process for sales and operations planning and was challenged by the wide range of approaches and data inputs from various stakeholders—suppliers; teams in the company's marketing, production, logistics, and supply sourcing functions; and customers who operated complex supply chains with thousands of retail store locations. High uncertainty in consumer demand led to constant shortages at some retail locations and excess inventory at others.

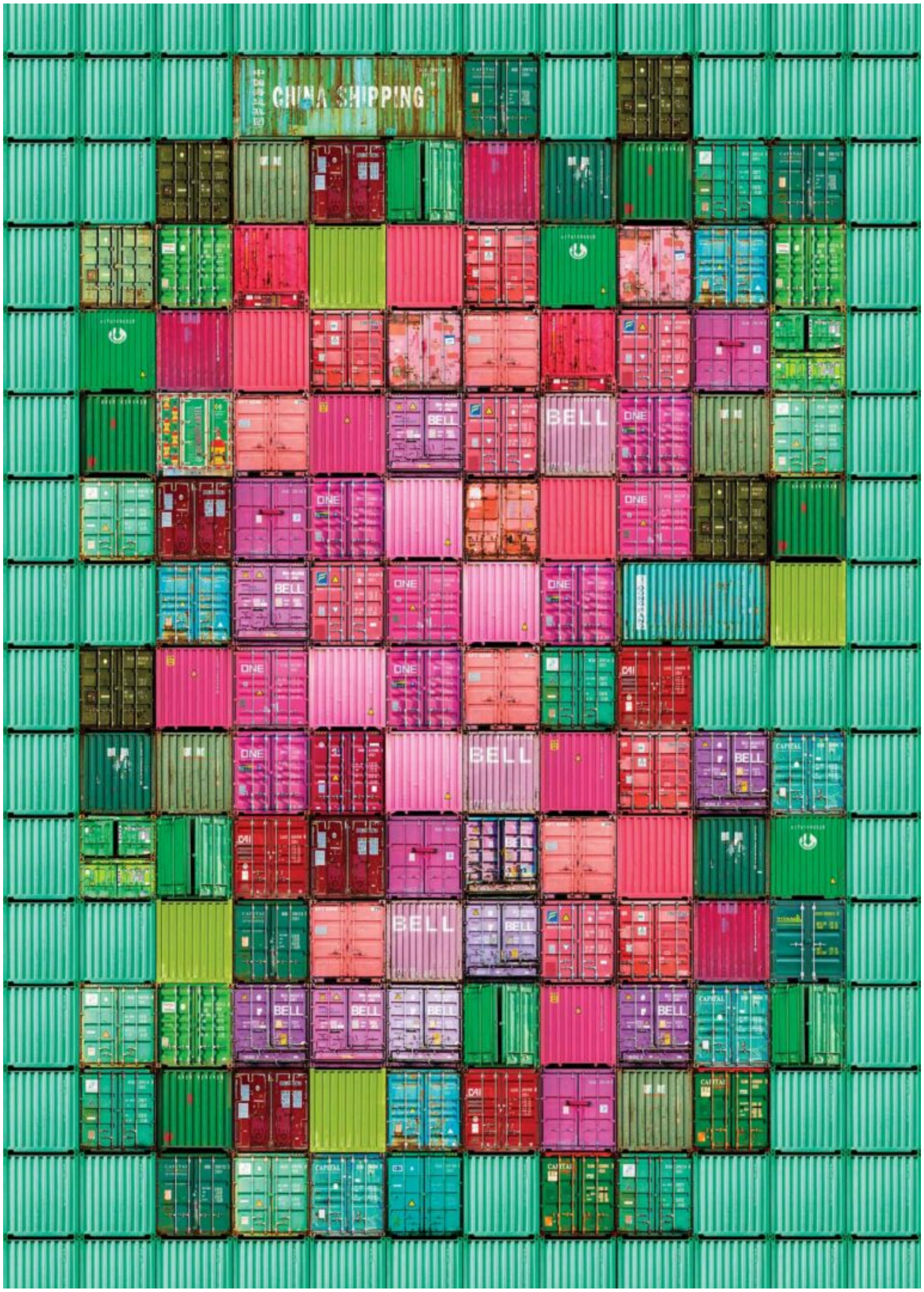
Why Agility and Resilience Remain Elusive

Companies across industries have failed to develop effective strategies for preparing their supply chains to withstand unforeseen disruptions. That's because of three significant shortcomings in existing planning methods: flawed, forecast-driven processes; data-related challenges; and ineffective scenario planning.

Forecast-driven planning. The most widely used approach for supply chain planning consists of two steps. In the *predict* step, demand forecasts are generated from historical data about sales and orders, information about economic conditions and competitors' actions, and subjective judgments. In the *optimize* step, those forecasts are fed into mathematical models of the supply chain network in order to generate final stocking decisions. This approach, known as predict-then-optimize (PTO), fails for various reasons.

1. There is no single forecast that all parties agree to use for decision-making. At the consumer electronics company, sales managers developed monthly forecasts of aggregate sales, along with forecasts of sales to each retail customer for the upcoming 12-month period. The inventory planning team independently created its own forecast of customer demand, but it did so weekly because inventory orders were typically placed once a week. To manage orders with the consumer electronics company and allocate inventory to its network of stores, each retail customer also developed its own forecast. These customer forecasts often differed significantly from those of the consumer electronics company. Moreover, all the forecasts were made at different points in time and were updated as new information about demand emerged. And the estimates of end-customer demand were often distorted because no one fully understood how marketing promotions implemented by the consumer electronics company or by retailers would affect sales. In situations like this, when multiple forecasts are available, each with its own errors, it is not clear which is the ideal one to use to optimize supply-chain-planning decisions.

2. The objectives of the various stakeholder groups in the planning process are not aligned, which leads to biased and suboptimal decisions. At the consumer electronics company, the sales team typically tended to overestimate forecasts to ensure supply, while inventory teams tended to deflate forecasts to ensure minimal obsolescence. The finance team cared primarily about obsolescence costs associated with unsold inventory. Senior management, of course, cared most about meeting quarterly revenue targets and having minimal capital tied up in inventory. The company's suppliers, for their part, faced production-capacity constraints and the competing demands of multiple customers; as a result, the company was often at the mercy of its suppliers' decisions. Meanwhile, to guard against the danger of receiving insufficient quantities of needed electronics products, retailers often inflated their orders.



Belgian photographer Didier Engels photographs docks, cargo hulls, and shipping containers throughout Europe. For these images he took aerial shots of colorful containers and manipulated them to create striking compositions.

Misaligned aims were an issue for the semiconductor equipment maker too. Inventory managers often kept too few parts in stock, hoping to minimize the overall investment in inventory and avoid costly write-offs of unsold items. That was because senior management provided targets for invested capital and write-offs without a clear understanding of the resulting potential for lost revenue.

3. The methods for deciding how to optimize inventories are flawed. The conventional models widely used today don't capture all the intricacies of supply chain dynamics. They generally incorporate overly simplistic algorithms for the optimization step in PTO. And even if advanced "optimizer" tools are used, any errors in forecasting, model formulation, and optimization will compound and lead to poor results.

An additional challenge in planning is that even if perfect forecasts were available, they would not suffice. Decisions about supply orders and shipments to various locations must also consider a variety of factors that will affect key financial and operational performance indicators. Those factors include constraints on resources (such as production, shipping, and storage capacity), customer-specific needs, differences in profitability by customers and regions, and both local and global service-level requirements.

Data challenges. It is now feasible to maintain and access vast amounts of data about store-level sales transactions in almost real time. Unfortunately, for companies with global supply chains, accessing and consolidating that data remains a mega challenge.

Consider the consumer electronics company. Its supply chain data was widely dispersed among various actors both within and outside the organization. Internally, sales teams maintained forecast information, marketing people handled promotions-related information, the supply planning teams managed inventory information, the finance department had responsibility for financial information, and so on. Externally, data on store sales and on shipments from customers' distribution centers to stores was under the purview of those retailers. Making matters worse, the data also resided in a variety of formats and repositories—partly in Excel files, partly in Tableau tables, partly in text form in emails, and so on. In some instances, data was maintained at an aggregate level rather than with the specificity needed for decision-making. Because of the scattered nature of the information, there was simply no mechanism for leveraging granular transaction data to support the analysis needed to drive efficient replenishment decisions. Moreover, it was impossible to analyze the impact of various external factors (such as business cycles) and competitive factors (such as the introduction and pricing of new products from rival firms) on the end-to-end supply chain because no single representation of the entire chain existed.

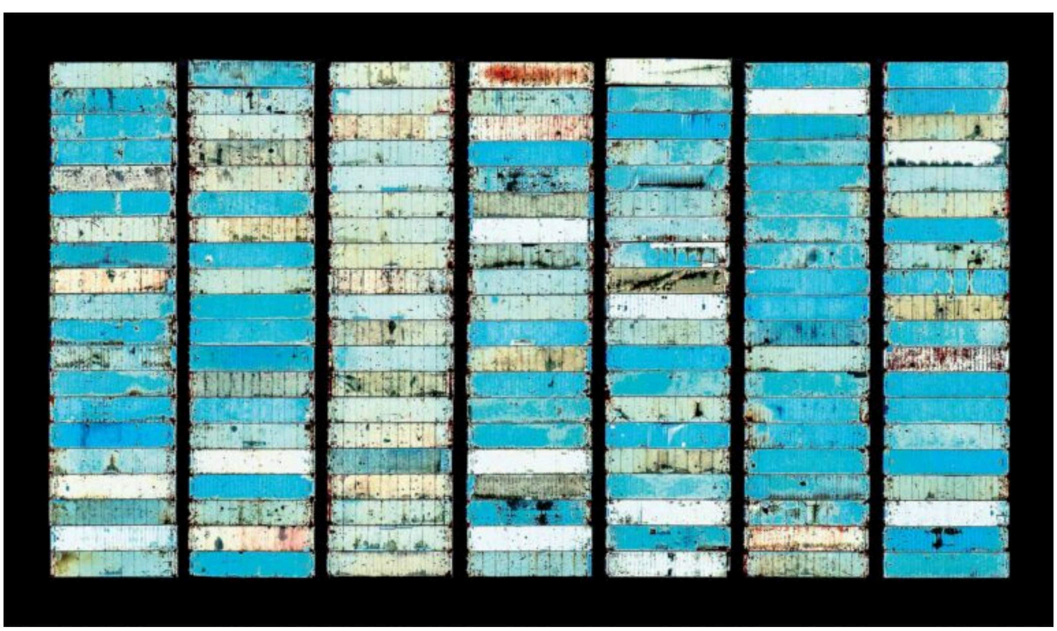
Ineffective scenario planning. A first step in developing strategies for making a supply chain more agile (able to respond to disruptions) and resilient (capable of bouncing back) is articulating future scenarios corresponding to the many risks that can affect supply chains. Black-swan events—rare, high-impact disruptions, such as the blockage of the Suez Canal by a container ship in 2021—are virtually impossible to predict. Other threats that can have a major effect on supply chains—such as the Russia-Ukraine war—can be foreseen, but the likelihood of their occurring can be difficult to ascertain.

Although many companies have started incorporating scenario planning into their supply-chain-planning processes, such analysis often lacks sufficient detail to be useful. For instance, it is not enough to know that a company's overall sales can drop in the event of a war that constrains a key supplier's operations. What is needed is a deep understanding of the magnitude and timing of the impact on each product, customer, and region so that appropriate strategies for ameliorating it can be developed. Such understanding is also important for creating bespoke supply chains—those differentiated by geography, products, and customers—rather than using a one-size-fits-all strategy. Unfortunately, absent a mechanism that can offer granular insights, planning exercises often lead to subpar outcomes.

A New Paradigm

Our approach, optimal machine learning, overcomes the significant shortcomings in existing supply-chain-planning methods. It has three key components: a decision-support engine, a digital twin, and an end-to-end data architecture.

Decision-support engine. OML replaces the forecast-based, two-step planning process with a one-step process that connects input data directly to supply chain decisions. The OML methodology incorporates historical data about drivers of demand throughout the supply chain (for example, actual sales transactions and factors that influence them, such as prices, promotions, and the size and composition of the customer base) and drivers of supply (such as supplier inventories, lead times, capacity constraints, and transportation delays). The engine determines the relationship between these supply-and-demand data elements and supply chain decisions about such things as production quantities, replenishment orders, inventory stocking levels, and shipments of inventory between locations in a way that optimizes key performance indicators (KPIs). KPIs might include metrics related to overall profit; revenue or profit by region, customer, or product category; product availability; inventory turns; time needed to recover from a disruption; and so on. Calculations can be recomputed as soon as new data is available or if any changes are made to the supplier network.



Our implementations of OML at both the semiconductor equipment maker and the consumer electronics company leveraged the power of cloud computing to run the decision-support engine. Both produced decision recommendations in minutes, despite the large size of the respective supply-chain networks. Our approach also can incorporate business constraints specified by managers (for instance, “My budget is X” or “I can produce no more than Y units” or “The availability of my product has to be greater than Z%”). A lot of traditional machine-learning approaches ignore such limitations. As a result, the solution has to be “fixed” after the fact, which is not ideal. OML does not face that problem.

Digital twin. A key requirement for the OML decision-support engine to work is a detailed digital representation of the entire supply-chain network, all material flows, and the decision-making processes of all involved parties. By modeling the performance of the supply chain in a highly detailed manner at the granular level of each transaction, the digital twin allows supply chain planners to quantify accurate KPIs for both past and future demand-and-supply scenarios.

Such a digital representation has two advantages. First, its calculations of the impact of historical decisions can be compared with the actual results computed by existing business systems, which validates its ability to measure performance accurately and thereby gives management confidence in the system. Second, it can be used to test the effects of alternative supply-chain scenarios or strategies on KPIs. For example, what if the shipment lead times double because of slowdowns in a specific shipping lane? Or how would the disruption of a distribution center affect revenue? Or what would be the best way to reroute shipments to stores served by that distribution center? Together, these two capabilities—historical analysis and predictive power—make it possible for managers to accurately evaluate risk-mitigation strategies, such as alternative sources of supply and transportation, alternative stocking locations, revised production schedules, and the use of product substitutes. As a result, managers can make better decisions.

End-to-end data architecture. The OML decision-support engine and the digital twin require a data storage system that works in conjunction with all existing database-management systems throughout the supply chain (those for the company's operations and those of suppliers, distributors, and customers). The storage system should be able to pool data across teams, locations, and products and make it possible to update and access that information in near real time.

The architecture specifies the various data elements to be included, their storage format and organization, linkages between them, and the frequency with which they will be updated. The choice of data to be incorporated should be driven by the analytical representation of the supply chain in the digital twin and should consider input from the managers who best understand the potential drivers of decisions. In other words, the OML approach entails the collection and storage of data that reflects the supply chain structure. We recommend using a graph database that houses the data at nodes in the supply chain (for example, retail stores or wholesale locations) and captures important relationships (for example, which wholesale location supplies which retail stores). This allows meaningful visualization of data and metrics by location, customer, product, or time. It also enables supply chain planners to measure performance for multiple metrics related to cost and service and to identify instances when the observed values fall outside the expected range and require further analysis.

It is important for senior managers to ensure that all parties trust the recommendations that come out of

the planning system and don't feel the need to review them.

This approach is in stark contrast to one that uses all available data, regardless of its business relevance. Some companies have chosen to "throw everything in" while developing machine learning models in the belief that the models will figure out what data matters most and then weight it appropriately. The problem with that approach is that the model is perceived as a black box, and when decision-makers can't understand why it produced the results it did, they often don't trust it.

Driving Results

At the semiconductor equipment company, the OML methodology was used to determine specific inventory policies that could lead to higher service levels (the fill rate) at a lower cost. (In any supply chain, higher levels of product availability require greater investments in inventory, which increases costs.) Prior to the deployment of OML, the company's legacy planning system could maintain a fill rate of about 77% with inventory investments of a little over $135 million. When managers used their expertise to override the system's recommendations, they were able to increase the fill rate to about 81% with a slightly higher inventory investment. The OML system gave the company the choice of achieving this higher fill rate while spending nearly $20 million less on inventory, or increasing the fill rate to nearly 85% while keeping the inventory investment at about $135 million. Moreover, the system freed up managers to focus on more-strategic issues. The visualization capabilities of the digital twin helped them understand exactly what needed to be changed and why, which increased their willingness to accept the recommendations. For instance, a key insight was that OML's decision rule considered the number of existing product installations along with new ones and captured their effects on inventory deployment decisions—linkages that had not been factored in previously.

In the case of the consumer electronics company, the OML model revealed glaring deficiencies in how inventory was being managed. For example, the distribution center that served the region with the highest sales volume had historically stocked the least amount of inventory. That led to frequent shortages at stores it supported, as well as expedited shipments to the center. The OML model correctly identified this discrepancy and recommended the optimal inventory-stocking policy for that location. Our analysis also revealed that while the average level of product availability at most retail stores was acceptable, a significant number of stores experienced severe shortages. It is important to note that while the OML model captured the entire network down to retail locations, all decisions about inventory deployment at customer distribution centers and retail stores were made by the customers, not the consumer electronics company. However, thanks to the model's insights into inventory deployment, the company was in a stronger position to influence its customers' inventory decisions. The key enabler in the conversation was an easy-to-interpret visualization of the model's decision recommendations that showed exactly which drivers were responsible for them. Such visualization is critical to securing the buy-in of all stakeholders.

Succeeding with the New Approach

For an analytics-enabled tool like OML to be helpful to companies in building agility and resilience in their supply chains, it must be accompanied by an appropriate organizational structure, personnel with the right skills, changes in the planning process, and a detailed understanding of the potential and pitfalls of machine learning.

First, senior executives should ensure that the planning team, from the outset, includes a broad range of internal and external stakeholders. Defining the OML objective, capturing business constraints, and identifying KPIs and relevant data will require functional experts and operations research analysts with sufficient experience in modeling supply-chain-management problems. This means that the team should include people from marketing, sales, finance, supply chain logistics, production, and IT; data scientists and analytics experts are also crucial. If necessary, the internal expertise of an organization should be supplemented by outside consultants and academic experts. Some companies delegate machine learning projects solely to data scientists. That is a mistake, because data scientists usually have minimal familiarity with the supply chain domain.

Second, OML requires companies to identify, assemble, and access the required data inputs from multiple sources and to verify the outputs generated by the machine learning model. The former requires computer scientists who can construct the necessary end-to-end data architecture using modern database systems, while the latter requires supply chain planners and managers with sufficient domain knowledge and experience.

Third, the sales and operations planning (S&OP) process, in which members of the planning team come together to develop sales, production, and inventory plans, must be redesigned to leverage the agility that OML can offer through its speed and depth of analysis. It is common for the S&OP cycle to be one month long because it takes that long to process information from the previous cycle and reach consensus on actions for the next one. But companies then struggle to respond promptly to disruptions in supply or logistics and to the shifts in demand that occur constantly. What they need to do is replace the typical monthlong S&OP cycles with faster and more-responsive ones. That will require investments in systems and processes that allow supply chains to react with speed, such as cloud-based state-of-the-art solver software systems that can generate solutions within a reasonable time frame (minutes, say, for a particular scenario or policy analysis). Companies thus need expertise in optimization and cloud computing, which they can get from software and cloud solutions firms, respectively.

Finally, the planning team, in consultation with senior management, should establish the KPIs that will ultimately drive decisions within the supply chain. These might include fill rates by customer, product, and region; costs and profit margins; and inventory turns. Understanding the conflicts between internal and external stakeholders and securing everyone's buy-in is important in tackling this step. By providing a detailed analysis of the implications of various decisions for each stakeholder, OML can help. Typically, decisions affecting overall supply-chain performance are made by multiple stakeholders. It would be best if the performance metrics for each stakeholder were transparent and accepted by all parties. Our OML approach allows the model to be run iteratively until it finds a solution that is mutually agreeable. Ultimately, it is important for senior managers to ensure that all parties trust the recommendations that come out of the planning system and don't feel the need to review them.

...

OML allows companies to base decisions on historical and current supply-and-demand information rather than just more-accurate forecasts. It gives them a tool that can help them reduce costs and increase revenues, profits, and customer satisfaction. It enables them to test strategies for mitigating risks, making it easier to choose the best ones. By doing all these things, it offers a way to build more-flexible, more-resilient, higher-performing supply chains.

A version of this article appeared in the March–April 2024 issue of *Harvard Business Review*.

Narendra Agrawal is the Benjamin and Mae Swig Professor of Information Systems and Analytics at Santa Clara University's Leavey School of Business.

Morris A. Cohen is the Panasonic Professor Emeritus of Manufacturing & Logistics at the University of Pennsylvania's Wharton School. He is also the founder of AD3 Analytics, a start-up that developed the OML methodology for supply chain management.

Rohan Deshpande is a machine learning scientist at Cerebras Systems and a former chief technology officer at AD3 Analytics.

Vinayak Deshpande is the Mann Family Distinguished Professor of Operations at the

University of North Carolina’s Kenan-Flagler Business School.