The analytics-enabled collections model

How leading institutions are using the power of advanced analytics and machine learning to transform collections and generate real value quickly.

Ignacio Crespo and Arvind Govindarajan

The global credit environment absorbed the effects of the financial crisis at varying speeds from market to market. In some places, loss rates have remained relatively high since 2008–09; in others, the past decade has been one of steady improvement, with tapering losses that have only recently begun to climb again. In the expanding markets, lenders increased their risk exposure, issuing new products designed around easier underwriting guidelines. Little attention was paid to maintaining or improving collections capabilities. As debt loads rise, however, institutions in these markets are beginning to rebuild collections staff and skills that eroded in the previous period. Meanwhile, in the more stressed markets, the need for more efficient and effective collections operations is likewise becoming a priority.

The need to renew collections operations provides lenders with an ideal occasion to build in new technologies and approaches that were unavailable when the financial crisis hit. The most important advances in collections are being enabled by advanced analytics and machine learning. These powerful digital innovations are transforming collections operations, helping to improve performance at a lower cost. Better criteria for customer segmentation and more effective contact strategies are being developed. Individual collector performance is being improved with better credit-management information and other tools. Contact can be managed through an array of channels, some allowing customers a greater sense of control over their finances. Loss-forecasting strategies can also be made more accurate and predelinquency outreach made more effective with enhanced financial tools and mobile apps.

Some of the most significant advances brought about by advanced analytics and machine learning are in customer segmentation, which is becoming much more sophisticated and productive. Better segmentation—including innovative behavioral segmentation, discussed in detail in an accompanying article—is providing the basis for more effective

collections processes and strategy. The improvements affect the complete collections agenda, beginning with the prevention and management of bad debt and extending through to internal and external account resolution.

A next-generation collections model

In traditional collections processes, banks segregate customers into a few simple risk categories, based either on delinquency buckets or on simple analytics, and assign customer-service teams accordingly. Low-risk customers are usually given to newer collections agents based on availability; the agents follow standardized scripts without being asked to evaluate customer behavior. Agents with moderate experience, training, and skills are assigned, again based on availability, to medium-risk customers. These agents also follow a standardized script but are trained to assess customer behavior based on ability and willingness to pay. High-risk customers are assigned to the most skilled agents, who own their accounts and use less standardized approaches to develop assessments of customer behavior. Contact strategies and treatment offerings are fairly varied across the risk categories.

By using advanced analytics and applying machine-learning algorithms, banks can move to a deeper, more nuanced understanding of their at-risk customers. With this more complex picture, customers can be classified into microsegments and more targeted—and effective—interventions can be designed for them (Exhibit 1).

Using analytics in the new model

Analytics-based customer segmentation is at the center of the next-generation collections model. The transformed collections model will allow lenders to move away from decision making based on static classifications, whether these are standard delinquency stages or simple risk scores. Early identification of self-cure customers will be one benefit. Another will be an approach based on value at

Exhibit 1

Advanced analytics and machine learning can classify customers into microsegments for more targeted interventions.

Customer type	True low-risk	Absent-minded	Dialer-based	True high-touch	Unable to cure
Targeted intervention	Use least-experienced agents provided with set scripts	Ignore or use interactive voice message (segment will probably self-cure)	Matching agents to customers and live prompts to agents to modify scripts	Focus on customers able to pay and at high risk of not paying	Offer debt-restructuring settlements early for those truly underwater
Impact	Agent–client conversation guided by on-screen prompts based on probability of breaking promises	10% time savings allows agents to be reassigned to more difficult customers and specific campaigns	Can lead to increased "connection" and higher likelihood of paying	Added focus addresses higher probability of default rates in this segment	Significant increase in restructuring and settlements enhances chance of collecting at least part of debt

risk, rather than blanket decisions based on standardized criteria. The aspiration is to have every customer as a "segment of one" with customized treatments.

Leaders taking the analytics-based actions that define the new model have already begun to realize gains in efficiency and effectiveness. One European bank automated 90 percent of communications with clients by developing two advanced-analytics models using machine-learning algorithms. A binary model identifies self-curers and non-self-curers, and a multiclass model recommends collections strategies for the non-self-curers, including soft measures, restructuring, or workouts. The models use around 800 variables, including client demographics and information on overdrafts, client transactions, contracts, and collaterals. The bank has realized more than 30 percent in savings with no loss in operational performance.

Another European bank set out to develop a top-notch recovery process using advanced analytics. The goals were to minimize the number of clients falling 90 or more days' past due while maximizing the economic impact of exits, focusing on retail and small-and-medium enterprise portfolios. As the bank gained a deeper understanding of its nonperforming loans, it was able immediately to address certain borrowers (such as recurring defaulting clients) with effective initiatives. Other groups of clients were identified, and exit strategies based on economic value were developed for each group. The results are compelling. The bank reduced its 90-day-or-more portfolio by more than €100 million, with €50 million in fewer past-due entries and the remainder in exit acceleration. Moreover, a reduction of 10 percent in past-due volumes was achieved across the board, worth around €300 million less in past-due exposure.

A leading North American bank has rolled out a number of machine-learning models that improve the estimation of customer risk, identifying customers with a high propensity to self-cure as well as those suitable for early offers. These models have so far enabled the bank to save $25 million on a $1 billion portfolio.

Most banks can achieve results of this magnitude by introducing an analytics-based solution quickly and then making needed improvements as they go. Value can be gained in almost all of the key areas in the collections environment.

- *Early self-cure identification.* Some banks use rudimentary heuristics (rules of thumb) or simple models to identify self-cure customers, while others have adopted simple self-cure models with limited variables. The new self-cure model based on machine learning and big data can save collectors a lot of time. By using many variables to better identify self-cure accounts, banks can increase collector capacity by 5 to 10 percent, allowing agents to be reassigned to more complex collections cases.

- *Value-at-risk assessment.* While many banks use time in delinquency as the primary measure of default risk, some lenders are taking a more sophisticated approach, building a risk model to determine value at risk. Many of these are simple trees and logistic regressions, however, with limited data. Leaders are moving to a future state in which models project conditional probability rather than assign customers single risk scores. The conditional score is dependent on a range of tailored approaches to customer contact and engagement: every borrower has several scores depending on the contact strategy and offer. Lenders would then use the strategy and offer that optimizes recoveries. The approach better calibrates the intensity of contact with each account, thus optimizing resources. A next-generation value-at-risk assessment can further reduce charge-offs by 5 to 15 percent depending

on maturity of current operations, analytics, and availability of data.

- *Cure assessment versus pre-charge-off offers.* At most banks, agents determine whether a customer will cure or will need an offer of some sort; some banks have heuristic rules for agents to follow. The new approach is to use models that ascertain a customer's ability and willingness to pay and gauge whether the better path is a cure or an offer. Banks can resegment delinquent accounts to improve their decisions to offer early settlement, an approach that increases the uptake of offers while reducing charge-offs by 10 to 20 percent.

- *Optimizing pre-charge-off offers.* Banks are currently using rules or simple analytics to create offers for customers, often without determining the likelihood that they will accept. Models will predict the best offer, optimized for the needs of the bank and the customer. Banks can change the prompt, adjusting loan characteristics and offerings to those most likely to reduce charge-offs, including re-amortizing the term or interest rate, consolidating loans, or settling. Making the right offer early, before accounts enter late-stage delinquency, can improve acceptance rates.

- *Post-charge-off decision.* Most banks use simple models or heuristics to determine which agencies to send accounts to, and at what price. To refine these decisions, models will determine the best agency for each account and tailor prices accordingly. The model will also determine the optimal pricing segmentation for third-party agencies and identify the accounts to retain in-house longer (based on products retained with the bank, for example). The strategic use of third parties can help with accounts that cannot be cured internally.

Integrated analytics models

Lenders at the forefront of the analytics transformation are assembling masses of data from many kinds of

sources and developing different models to serve collections goals. The data sources can include customer demographics, collections and account activity, and risk ratings. The most sophisticated lenders are creating "synthetic" variables from the raw data to further enrich their data. Machine learning helps identify markers for high-risk accounts from such variables as cash-flow status, ownership of banking products, collections history, and banking and investment balances. By using so many inputs from many different systems, lenders can dramatically improve model accuracy, lower charge-off losses, and increase recovered amounts. Two separate institutions recently adopted similar approaches using more than 100 variables to support numerous machine-learning models. These issuers used

machine learning to identify the optimal treatment and contact strategy for each delinquent account; deployed the solution inside the existing collections workflow environment; and trained collectors to use the system and collect additional data to improve model performance. The initiatives were up and running in about four months. (Exhibit 2).

Contact strategies and treatment approaches

Institutions adopting the most analytics-forward approaches have been intensifying the development of new treatment and contact strategies, expanding the limits of digital capabilities. By applying advanced analytics and machine learning, banks can identify the most promising contact channels, while also developing digital channels to define innovative and

Exhibit 2

Two major issuers adopted a new approach using more than 100 variables and machine learning to accelerate development of treatment and contact strategies.

Integrated data sources on client behavior	Multiple machine-learning models used to identify features of high-risk accounts	Implemented in the existing collections environment
Customer demographics	Low cash flow	Contact-center interface
Collections activity	More banking products	Banking machine (automated touchpoint)
Account activity	Previously in collections	Contact customer via call or text (automated touchpoint)
Payments	High total balance across products	Interactive voice response (automated touchpoint)
Risk ratings	Low investment balance	
• 140 inputs across 15 systems describe profiles of each client on each day in collections	• 30–40% improvement in model accuracy versus previously existing models	• Implemented in 12–16 weeks

Impact: 4–5% lower net charge-off losses

regulatory-compliant contact strategies. The same digital channels can be used to build awareness of payment options.

- **Websites** can display messages and repayment options as soon as customers log in, increasing awareness and providing opportunities for early delinquency reduction.

- **Messenger and chat.** Where legally permissible, collectors can contact customers and negotiate payment options with chat functionality and free messenger applications (such as WhatsApp).

- **Mobile apps.** Build collections functionality into the mobile app, reminding customers in early delinquency stages to pay and offering payment options.

- **Virtual agent.** Create capacity by developing virtual agent functionality to call customers in early delinquency stages.

- **Voice response unit.** Enhance current voice-response capability, offering basic repayment options when customers call, which frees collector capacity.

Most banks use heuristics to establish the best times to call. Usually, however, agents are inadequately supported on questions of which channel to use, when to use it, and what the message should be. Advanced models can project a full channel strategy, including channel usage, timing, and messaging. Banks will be able to control contact down to the hour and minute, as well as the sequence of communications—including voice, text, email, letter, and interactive voice message. The approach is developed that maximizes the right-party contact rate (RPC) and influences customer behavior to prioritize payment. Such optimal contact sequencing can increase success in early stages of delinquency.

The analytics focus on the front line

Leading companies in many sectors—digital giants, healthcare providers, retailers, and manufacturers—are using data and analytics to develop a workforce optimized to business goals. Analytics is now the source of improved performance in realizing talent strategies as well as a means for linking talent strategy to business needs. Presently, recruiting and retention are often based on legacy processes, including résumé screening and interviews; retention is based solely on performance. Analytics can improve hiring, finding agents with affinities to the most valuable at-risk segments, as well as help identify collectors at risk of leaving. Companies are using machine-learning algorithms to screen résumés and to determine the value of external hiring compared with internal promotion. One global digital company used analytics to create a checklist that boosted onboarding speed by 15 percent. The algorithms, it should be stressed, are not replacing human judgment but are rather providing a deeper fact base for the exercise of informed judgment.

Companies are also using algorithms to uncover the bottom-line impact of employee engagement and to drive deeper engagement across the organization. In collections, where retention of talent is a recurring issue, people analytics can be used to find the drivers of performance, including personality profiles and risk factors for low performance and engagement. By identifying individuals most at risk of leaving, for example, banks can take responsive measures to optimize their talent pool for sustained performance improvement.

Machine learning and nontraditional data have become the new frontier in collections-decision support. Audio analytics, for example, is now an important tool for understanding frontline effectiveness. By allowing algorithms to work through thousands of conversations, banks can discover the most productive and engaging approaches. With hypotheses informed by insights from the field of behavioral science, banks are also using machine

learning to diagnose and neutralize the biases that affect collector and customer decision making. At the same time, the machine-learning approach is enabling automation of larger classes of decisions. By giving agents more prescriptive decision support in certain situations, including a wider range of set script elements and narrower parameters for negotiations, banks can free capacity and redirect resources toward the most valuable accounts. In this vein, one card issuer achieved dramatic improvements in the rate of promises kept in their high-risk segment by using an approach enabled by data and analytics to script elements, including behavioral insights (Exhibit 3).

Behavioral pairing and agent coaching

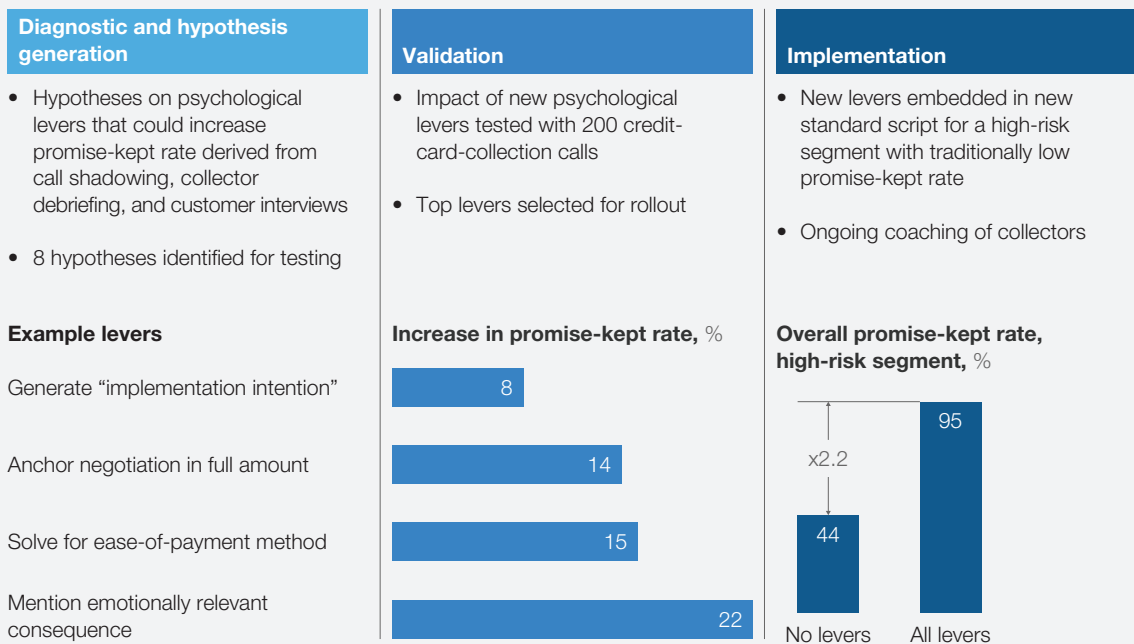Many banks do not apply agent–customer pairing uniformly or deliberately. When it is applied, high-risk customers are usually given to experienced, high-performing collectors, while low-risk customers are assigned to new collectors. Analytics-aided pairing helps match collectors and customers who have similar personal profiles. By smarter pairing—matching delinquent clients with the agent expected to be most effective—outcomes can be improved and call times reduced. As for coaching, this has often occurred in training sessions, huddles, and call monitoring by managers. Analytics-aided coaching permits real-time feedback and analysis in live phone calls.

Breaking through artificial barriers to transformation

Most banks understand that analytics and digital automation will transform their collections operations. Some have been reluctant to get started, however, in light of following persisting myths about the new technologies:

Exhibit 3

One card issuer systematically identified and implemented 'assertive' script elements, doubling the promise-kept rate to 95 percent.

Diagnostic and hypothesis generation	Validation	Implementation
• Hypotheses on psychological levers that could increase promise-kept rate derived from call shadowing, collector debriefing, and customer interviews • 8 hypotheses identified for testing	• Impact of new psychological levers tested with 200 credit-card-collection calls • Top levers selected for rollout	• New levers embedded in new standard script for a high-risk segment with traditionally low promise-kept rate • Ongoing coaching of collectors

Example levers / **Increase in promise-kept rate,** % / **Overall promise-kept rate, high-risk segment,** %

Example levers	Increase in promise-kept rate, %
Generate "implementation intention"	8
Anchor negotiation in full amount	14
Solve for ease-of-payment method	15
Mention emotionally relevant consequence	22

Overall promise-kept rate, high-risk segment, %

No levers	44
All levers	95

x2.2

- "Sophisticated data infrastructure is a prerequisite." While this is an advantage, it can be developed over time. The truth is that banks can build value-enhancing collections models with available data. As the data are improved, the models can be updated accordingly.

- "Both the collections front line and the digital infrastructure need to be in place before analytics models can be implemented." Actually, models can be implemented using legacy infrastructure, and the value they generate can be used to invest in the needed infrastructure improvements.

- "The development and implementation of models take a long time." Banks can get started using agile model development with minimum viable products subject to continuous improvement. Without rapid iteration and deployment of models, value is left unrealized.

- "Given compliance and regulatory issues, models are too opaque to use." Banks can select among a range of modeling techniques with different levels of transparency. They can balance demands for transparency and performance by choosing the most appropriate algorithms.

- "Success depends on nontraditional data." For most collections applications, banks' internal data can provide the majority of the gains from advanced analytics. Banks can begin by utilizing all internal data and supplement with external data subsequently as needed.

- "Regulations and compliance negate many of the benefits of advanced analytics and machine learning." A number of banks in highly regulated jurisdictions have already successfully deployed machine learning. Indeed, machine learning can improve compliance by better matching the right treatment with the right customer and avoiding biases.

None of these myths should prevent banks from beginning the analytics-enabled transformation of their collections operations. There is no perfect way to start a transformation—some of the implementation might even be messy at first. The essentials of the analytics transformation in collections are clear, however. First, set a long-term vision, but also a path toward it that generates value continuously. Second, work in an agile manner, with teams from all dimensions of the transformation. Focus on implementing working models from day one, avoiding an overly complex academic approach. Use synthetic variables to enhance model performance, and continuously experiment with strategies to generate additional data for the next generation of models.

◼ ◼ ◼

The next-generation collections environment will be built around advanced analytics and machine learning. These approaches will help lending institutions meet the new delinquency challenges that market analysts predict are on the horizon. The transformation of collections has in fact already begun, as leading institutions assemble the data and develop algorithms to attain improvements in their existing collections context within a few months' time. These leaders are showing the way by applying the new approaches and making improvements as they go. And they are already generating bottom-line results. ◼

Ignacio Crespo is a partner in McKinsey's Madrid office, and **Arvind Govindarajan** is a partner in the Boston office.