When Is Big Data the Way to Customer Centricity?

Next-generation analytic learning finds critical insights in an ocean of false clues

By Dr. Andrew Jennings, FICO Chief Analytics Officer and Head of FICO Labs

Two trends challenging business thinking today—Big Data and customer centricity—seem, at first, to be antithetical. Driving decisions from more and more data raises the specter of dehumanizing business interactions. But the real value of Big Data for business is the opportunity to learn about our customers at such depth and speed that we can truly put them at center stage.

Still, answers to the most important questions—How is this customer likely to respond to this action? What new needs can we anticipate? What does this changing behavior mean?—aren't just there to be scooped up from Big Data. It's awash with noise, distractions, redundancies and useless correlations. How do we plunge our nets among these false clues and come up with the accurate, reliable high-ROI insights we need to make our businesses customer-centric?

In this paper, I'll discuss next-generation analytic learning. By this I mean data-driven learning enabled by computing infrastructures and analytic techniques that make it practical to examine data of very high volumes, variety and velocity. When these techniques are systematically employed, Big Data yields what is sometimes called its “fourth V”—value.

Companies that become very good at next-generation learning will be able to orient their entire operation around their customers. They'll engage customers and build win-win relationships with such insight, innovation and efficacy that they'll be very difficult to dislodge as providers of choice.

I'll cover four analytic imperatives for next-generation learning:

1. Design and automate smart experiments that enable causal prediction
2. Analyze and learn from customer behavior on the fly
3. Get the machine learning / human expertise balance right
4. Turn every customer touch into an opportunity for more service and learning
Businesses in many industries across the globe are launching major initiatives in customer centricity. They may phrase the goal or define the term differently, but all are aiming for a higher level of performance driven by greater understanding of customers.

It’s no longer enough to have a 360-degree customer view linking data from all products and channels. Today we’re turning that inside out—to see the relationship from the customer’s point of view and operate from that perspective. To be strong competitors, we must now anticipate customer needs, understand transaction context, notice what’s changing and drive personalized treatments across all touch points within the window of opportunity.

Big Data can help or hinder us on the way to customer centricity. Today we have the means to capture and analyze much bigger quantities of data than ever before, and to make meaningful connections between different types of it. We can analyze data in-stream for real-time decisions. We can distribute analytic tasks in a massively parallel manner across many processor nodes, then algorithmically assemble their outputs into a single result. But is any of that helpful for achieving customer centricity?

It’s helpful only when we can systematically extract the most valuable analytic insights—causal relationships—from Big Data. These insights enable us to understand individual customer behavior and sensitivities, anticipate needs, and predict likely responses to offers and treatments. In some situations, we must find and act on such insights as data is streaming in. In others, we can use out-of-stream methods to dive deeply for them.

Big Data computing infrastructures are making it practical to employ automated machine learning algorithms for this purpose—but human expert oversight is essential to ensure results make business sense and are useful in operations. And, ultimately, whether any of these insights have any impact at all on customer centricity depends on how quickly we can pump them into operations so that they drive or inform every decision we make and every interaction we have with our customers.

These are essential capabilities for turning Big Data into an enabler for customer centricity. They’re the fundamentals of what I call “next-generation learning.” Next-generation learning starts with what you want to know about your customers—in other words, with business questions like “Which of my customers are most sensitive to discount coupons?”

Starting with the business questions helps you target the right data. This is the case for a national US bank with a CEO-mandated customer centricity initiative. FICO is advising the company on the highest priority data within its vast integrated data warehouse for answering key questions and deriving strong customer analytics to support this initiative.
When Is Big Data the Way to Customer Centricity?

I call this approach “next generation” because, as depicted in Figure 1, it elevates test-and-learn methods to a new level of efficacy. These improvements to the traditional champion-challenger method were not triggered by Big Data; they’ve evolved in response to increasingly granular customer treatments and rapidly changing customer behavior. Still, next-generation learning is certainly suited to the challenges and opportunities Big Data presents. It’s a systematic, highly efficient way of continuously advancing what we know about our customers and improving how we use those insights to interact with them.

It’s important to note that next-generation learning can be practiced by every company—not just those who’ve mastered champion-challenger. In fact, I’ll talk about how to get the most value from what you have right now, even if your data has “holes” and other imperfections, and your customer treatment strategies have been inconsistently applied.

Here are four fundamentals that, in my observation, distinguish companies beginning to make Big Data pay off for business:

1. Design and automate smart experiments that enable causal prediction

While the easiest relationships to find in Big Data are correlative (when A occurs, B also occurs), the most valuable for customer centricity are causal (A affects B). By finding and testing causal relationships (A affects B in this specific way), companies can better understand customer sensitivities to offer features such as price, brand status or flexible terms, and predict how individual customers are likely to respond to a specific offer or treatment.
FICO captures causal relationships in the action-effect models at the heart of our approach to improving decisions through decision modeling and optimization. (Figure 2 shows a simplified diagram.) Well-constructed action-effect models enable us to answer questions like: “How will profit and loss likely change if I give Jim a $2,000 credit line increase instead of a $1,000 increase?” or “Is a 20%-off coupon likely to affect Jane’s propensity to buy this camera in the next two weeks?”

Case in point: A major retailer wanted to understand how to target its discount coupons more profitably. The central question was: Which customers will buy only with the coupon and which will buy whether or not they receive the discount? To answer that question, we needed to find causal relationships between coupon treatment characteristics and buying behavior.

While the available historical data was big in volume, it was poor in causal insights. The retailer had done no deliberate experiments on coupon effects. Historically, its coupon distribution was determined by specific customer characteristics (e.g., past purchases, frequency, demographics). Similar customers, therefore, almost always received the same coupon. As a result, the data had very little of what analytic scientists call “common support,” but which we can simply think of as overlap. Overlapping different treatments across similar customers enables us to compare outcomes and see how the treatments affected them—essential for finding causal relationships.

If the company had randomly distributed coupons, the result would have been worse in terms of costs but better for learning. Random assignment creates lots of overlap in the data. (In fact, sometimes data you’d ordinarily regard as problematic, due to poorly directed, haphazard treatment assignment, can turn out to be a “natural experiment” you can leverage for analytic learning.)

This situation highlights a critical dynamic between customer centricity and Big Data learning. As our businesses become more customer-centric, we’re able to target treatments with finer granularity, right down to the individual customer level. But the more we target, the less overlap our data has and, therefore, the less opportunity it affords for learning about causal relationships.

To learn more, we need to introduce a certain amount of randomness into treatment assignment, while keeping testing costs in check. I’ll talk about new ways to do that—yielding faster learning than champion-challenger testing—in a moment. But first, let’s look at how to make the most from existing data.

In the case of this retailer, FICO found the answers the retailer needed using the data it had. One technique we employed was propensity scoring to locate useable overlap in the existing data (see Figure 3). Unlike the more
familiar application of such scores to predict likelihood a customer will buy, in this case we were predicting likelihood of a customer being assigned to a particular treatment. We found overlap in customers with similar propensity scores who had received different treatments, and within the overlap area were able to create matched samples suitable for identifying causal relationships.

One takeaway from this case study is that there may be value hiding in your existing data—in other words, more opportunity than you realize for learning about your customers. Another is that companies wanting to compete by understanding customers better need to deliberately enrich their data for learning. That means going beyond traditional champion-challenger to smart experiments that increase overlap for causal analysis.

Take a look at the center graph in Figure 4, which shows data generated by a typical champion-challenger contest. Customers receiving different treatments are indicated by the turquoise, blue and magenta dots. There’s more overlap in this graph than in the one on the left, which shows the results of determining treatments based strictly on customer characteristics.

The overlap with champion-challenger is limited, however, for two reasons. First, champion-challenger is randomly assigning customers to a strategy, not a treatment. The treatment they actually receive depends on how they fall through the nodes of the strategy decision tree to a particular end node—which is a deterministic process driven by business rules. Secondly, usually only 25% or so of the population is typically assigned to challenger strategies.

The chart on the right in Figure 4 shows the bigger overlaps we can achieve through designing smart next-generation learning experiments that support the discovery and modeling of causal relationships. We can use techniques, such as boundary-hugging test design, to expand overlap. This method actually tests treatments on individual customers rather than strategies.
It does this by algorithmically “flipping” treatments on a random basis for some customers. The flip probabilities are such that customers close to decision boundaries—that is, more similar to each other—are more likely to receive the treatment for the opposite side of the boundary. Customers farther away are less likely to be flipped, which keeps testing costs in check.

Moreover, we can automate the design of these next-generation challengers to accelerate learning cycles and balance testing cost. FICO’s patent-pending methods use machine-learning algorithms to rapidly generate potential challengers that produce high-overlap data and meet other requirements for supporting reliable estimation of causal effects. Because these requirements are met, there’s much less guesswork in the construction of action-effect models.

Automation also means companies can incorporate more models with strong causal predictiveness into their decisions while, at the same time, reducing the time business users spend building challenger decision trees. Instead, business users can think more about the extent of the learning experiment (e.g., define the range of scores and decision variables—price, limit, promotion offer, etc.—over which it will run).

In automated challenger generation, the algorithms constrain testing risk and cost by controlling the amount of deviation from the current champion strategy. Using simulation, companies can find their own “sweet spot,” balancing learning speed and investment for high ROI.

Figure 5 shows simulated results for overlap, incremental profit from strategy improvement (solid blue line) and testing cost (solid red line) of a moderately aggressive test design as it evolves toward the unknown optimal strategy (dotted green line), achieving 93% of optimal profit over five cycles. A more timid strategy costs less but achieves only 69% of optimal profit over the same number of cycles.

**Figure 5: Finding the “sweet spot” where learning investment pays off**
2. Analyze and learn from customer behavior on-the-fly

To operate in a customer-centric manner, businesses need to be able to respond to customer actions as they take place. In many situations, therefore, data analysis and data-driven decisioning must occur in real time or near real time, based at least partly on the stream of data coming in from mobile phones, ATMs, online activity, point-of-sale (POS) devices, sensors, etc.

The value of streaming analytics has been well-proven in credit card fraud management. Here, models detect unusual patterns of cardholder behavior and instantly generate a ranked score indicating how suspicious the transaction it is. These analytics not only drive real-time customer decisions, but can access and dynamically update profiles for individual cardholders, ATMs, POS devices and other entities involved in transactions.

Telecom companies have also used streaming analytics to prevent revenue losses and network outages by detecting unusual patterns of behavior in their billing and service delivery infrastructures. For one telecom company, in fact, FICO analytics examined 1.5 billion transactions a day, dynamically updating five different profiles (sending and receiving phone numbers, network nodes, routes and international interchanges). As in many high volume/velocity problems, historical data for training models was limited. Instead, we had to build models capable of learning about usual vs. unusual behavior from production data.

This requirement to build models that don't require supervised training with historical data is likely to become more common in the era of Big Data. Although vast amounts of data are being amassed, the dynamic nature of business today means it will not always fit modeling objectives and timeframes. As new ways of interacting with customers emerge and companies want to analyze data in innovative ways that produce a competitive advantage, relevant historical data may be limited, problematic or nonexistent. Where the data does exist, customer behavior may be changing so rapidly that traditionally trained models still need the ability to learn on the fly.

We can see the need emerging in banking, an increasingly dynamic industry. A South African bank, for instance, has seen a sudden rise in fraud on account-to-account transactions. With limited data at hand, the company turned to FICO for streaming self-learning analytics to solve the problem. More generally, we expect that first in Europe, and eventually in the US, this type of analytics will grow in demand as banks seek to garner more insights about their customers from activity going on in and through these core relationships, and build out profitable layers of business around them.

These streaming analytics are based on a FICO technology called self-calibrating outlier models. They compare the behavior of peers (e.g., similar customers, similar accounts, similar mobile connections, similar telecom switches) in order to detect outliers and score them for degree of deviance from the norm. The patented technology is critical to streaming analytic problems where the algorithms must continuously update, in real time, estimates of feature distributions so that detection of outliers is always based on the current distributions.

The first picture in Figure 6 (next page) shows the general idea for a single characteristic: high-value transactions per day. An individual in the outlier range for this characteristic could be perpetrating or experiencing fraud. On the other hand, she could be experiencing rising prosperity; outlier models can also point companies to opportunities to offer additional products and services to their customers.
When Is Big Data the Way to Customer Centricity?

„ insights

This picture, however, shows just a slice of what the model is doing, frozen at a moment in time. In practice, self-calibrating outlier models examine many customer characteristics in an instant, bringing context to unusual behavior. This complex analysis, based on real-time updates of all distributions of characteristic variables, also improves understanding of the meaning of outlier behavior.

The outlier zone is not static. As illustrated by the second Figure 6 picture, the model learns as normal behavior for the peer group changes over time. It’s self-calibrating because it dynamically readjusts the range of outlier values for each customer characteristic against this moving range of normalcy.

As such, this kind of model is an example of a good marriage between human expertise and machine learning. As with supervised models, the first step in building an unsupervised self-calibrating outlier model is decidedly human. Determining which customer characteristics are highly predictive and how to incorporate them in the model requires deep human understanding of the domain and problem to be solved. But while supervised models are trained with months of historical data to recognize normal and abnormal values for those characteristics, in self-calibrating models, machine learning takes over, inferring these values from the stream of transactions.

This approach is both powerful and flexible. Figure 7 depicts a multi-layered self-calibrating outlier model, a patent-pending technique FICO is implementing for an Asian bank. It uses parallel processing in a manner akin to a supervised neural network. In a neural network, multiple nodes of the model simultaneously look for nonlinear relationships across customer characteristics. Here the nodes are separate self-calibrating outlier models (light gray ovals), and the characteristic variables (blue, red and yellow ovals) hook up to them (gold lines) in different ways. Thus, each model has a slightly different view on the streaming data, which affects its outlier score.

The separate scores are then analytically combined (dark gray oval) into a single score, which now incorporates these multiple views, and therefore more context and nuance, increasing score accuracy and usefulness. These unsupervised analytic techniques are showing strong performance compared to supervised methods, and have the advantage that they continuously adapt to the production environment and customer behavior.

Streaming analytics, whether the models were built with supervised or unsupervised techniques, can be informed by other analytics. This is a very promising approach to further Big Data and customer centricity goals. That’s because it provides a practical way of combining the velocity of in-stream data analysis with other types of out-of-stream analytics that contribute insights from data volume and variety.

As an example, we can look at a current FICO service that provides fraud management systems all over the world with weekly updates of dynamic profiles on merchant activity. These profiles are generated in batch mode from huge volumes...
of transaction data. Moreover, by capturing the merchant view of transactions, they increase data variety, enabling the receiving systems to detect suspicious activity not visible from cardholder behavior patterns alone. The result is up to 15% better fraud detection.

We can expand and generalize the concept, as depicted in Figure 8. Rapidly deployed self-calibrating outlier models can be informed by other analytic insights. We can feed them with outputs from neural network models and the results of link analysis and semantic scorecards. We can leverage new types of databases (e.g., GraphDB, Triplestore) that don't require traditional data schemas and new ways (e.g., Map-Reduce) of finding, filtering and transforming data across massively distributed file systems. The flexibility of this architecture lends itself to a variety of analytic solutions for purposes such as cross-selling and pre-delinquency treatment.

3. Get the machine learning / human expertise balance right

In the two previous sections, I've discussed analytic methods that very effectively combine machine learning with human expertise. I want to underscore the need for this balance, since in the hype around Big Data, its importance is often underplayed.

Consider the development of a predictive scorecard. We can often boost performance by creating segmented scorecards that go beyond analyzing customer characteristics in an additive manner to examine complex interactions between them. But since there is an immense number of possible segmentation schemes, it’s unlikely the best scheme, producing optimal performance, will be found in a timely manner.

Machine learning speeds up the search by crunching through Big Data to test large numbers of characteristic interactions. For example, Tree Ensemble Models (TEMs), a type of machine-learning algorithm, have proved helpful in finding segmentation schemes that capture more complex customer behavior patterns with less danger of overfitting to “noise” (relationships specific to the development data and thus not generally reliable for making predictions on production data). In a laboratory study for a Chinese insurance company, FICO found that a TEM was nearly twice as effective at predicting auto insurance fraud as a traditionally developed model.

To make these insights useful in operations, however, requires human involvement. Analytic expertise is essential to compensate for biases and “holes” in the development data and for bridging the gap to production data, which will be different and will vary (often rapidly) over time. Moreover, instead of deploying the TEM itself—essentially a “black box” of hundreds of decision trees, difficult to understand, deploy and explain to regulators—FICO has innovated a method of transmuting TEM insights into a segmented scorecard. This technique, which we’re now using successfully for clients outside of the lab, has the advantage of greater transparency and more straightforward implementation. In addition, scorecards allow business experts to incorporate domain knowledge into predictions and customer treatments.
We can see the need for a similar partnership in the analysis of text and voice records. Businesses have collected a lot of this information—along with other forms of unstructured/semistructured data, it now accounts for some 90% of consumer data—but have so far analyzed almost none of it.

FICO research demonstrates that text analysis, while having some predictive value of its own, is extremely effective when folded into structured-data models. Initial steps include data cleansing and standardization (analyzing collector notes, we found 90% of the content consisted of abbreviations, codes, misspellings and garbled text). As depicted in Figure 9, we next extract text features and transform them into numerically based customer characteristics. A wide range of techniques can be used to accomplish this, from simple keyword flagging and indexing to more sophisticated methods like FICO’s patented context vector modeling, which examines clusters of words in order to understand meaning, and LDA, a new patent-pending FICO topic analysis method.

Because the number of characteristic candidates in text/voice can be quite huge, automation is very helpful. With Big Data technologies, we will have an increasing number of text mining methods available for crunching through unstructured data to find patterns. We can expect these to become quite commonly used.

The real opportunity for competitive advantage rests in human involvement—in how experts who understand the business problem apply automation to extract the most useful characteristics, and in how they incorporate text-derived numeric characteristics with the more traditional data in predictive models. A FICO study shows the potential: incorporating text characteristics from collector notes into a neural network model predicting likelihood of payment for delinquent accounts resulted in performance lift of 71% to 88% over the baseline model, depending on the sophistication of the analytics used.

Human expertise is also essential for making the most of text characteristics once they are in numerical form. FICO research has shown, for example, that insights from loan application text can lift the performance of a risk scorecard, but some of the same text characteristics could also be incorporated into propensity models targeting customers for cross-marketing.

More broadly, there’s potential to use text analytics (particularly methods examining context) to gain insight into customer attitudes and intentions, which have been difficult to analyze from structured data alone. More perceptive analytics could, for instance, help creditors across many industries target pre-delinquency treatments that minimize losses while preserving and nurturing valuable customer relationships.

1 IDC Digital Universe Study, sponsored by EMC, June 2011
4. Turn every customer touch into an opportunity for more service and learning

Promising analytics-driven decision strategies must be immediately deployable into operations in the form of business rules that automate or guide interactions with customers through all channels. (The top-down approach of starting from business questions and the transparency provided by the techniques I’ve been discussing helps with this process. The better you understand how you’ve constructed your answers, the more directly and reliably you can deploy them into operations.)

Streaming analytics and batch analytic results, deployed as an integral part of these strategies, work together. Entering into each interaction, the company knows who the customer is and has already made decisions about a range of appropriate treatments. During the interaction, the company makes additional intelligent decisions on the fly based on customer reactions and new data.

With these capabilities, businesses have the opportunity to transform customer experience. Take, for example, a retail banking customer, Jane. In the past, the bank’s ambition might have been to deliver relevant ads to Jane as she withdraws funds from an ATM—possibly her only point of contact with the company. Today, as depicted in Figure 10, the aim is to make the most of this opportunity to be of service to Jane, expand points of contact and keep learning more about her.

**Figure 10: Transforming the customer experience**

Jane withdraws funds at ATM from current account and savings.

- Streaming analytics detect outlier behavior patterns indicating short-term financial discomfort; transactional profile indicates likely external loan payments.
- Batch analytics (product propensity, risk score inputs to action-effect models for decision modeling/optimization) identify profitable loan offers (balancing response, take-up, revenue objectives with risk, capital constraints) for Jane.
- Channel propensity models identify best methods/times to contact Jane.

Bank sends Jane email and SMS saying it can offer a more beneficial loan product that will save her money.

Jane clicks on SMS, which takes her to a personalized web page.

At personalized web page, Jane selects one of several loan offers, interactively changing the term.

- Streaming analytics adjust loan details to Jane’s choices and, if necessary, invoke real-time optimization engine to negotiate with Jane.
- She can choose to immediately chat/speak with a customer service agent, who will be fully informed of all her information and online choices so far.
- Customer record/profile updated throughout dialogue.

Jane accepts an offer and is informed that she can come to a personalized web page any time to see the status of her loan—another touch point. She’s also offered a mobile app (bank-branded, preconfigured for her phone) to track her payments.

Jane downloads the app, opening another touch point.
When Is Big Data the Way to Customer Centricity?

The Insights white paper series provides briefings on research findings and product development directions from FICO. To subscribe, go to www.fico.com/insights.

How the leaders are doing it

Key capabilities for achieving customer centricity from Big Data

Here are key ways that FICO clients—including the leaders discussed in this paper—are using next-generation learning to leverage Big Data and make their businesses more customer-centric:

- **Predict customer reactions to offers and treatments.** Companies in retail, financial services and insurance have achieved substantial performance gains (10–40%) working with FICO to identify the strongest predictors of customer behavior and incorporate these into data-driven decision strategies. Powerful techniques used to create competitive advantage include action-effect modeling (causal relationships); decision modeling (balancing multiple objectives); and strategic-level, customer-level and product-level optimization (finding the best strategy within constraints) and simulation.

- **Add self-learning analytic techniques to adapt to changing customer behavior.** FICO pioneered several dynamic technologies that can be used flexibly in new model deployments and even “bolted onto” existing models. These include not only self-calibration techniques, but intelligent global profiles that update with new transactions, and adaptive models that adjust their own variable weights based on performance feedback data. Depending on how used, clients are achieving 5–40% performance improvements.

- **Use more streaming analytics to support real-time decisions during customer transactions.** No one has more experience in high-volume, high-velocity transactional analytics for customer decisions than FICO. Our streaming technology processes terabytes of transactions daily and is relied on by 17 of the top 20 credit card issuers worldwide to protect transactions in more than 2.6 billion active accounts.

- **Automate intelligent, multichannel communications.** Data-driven communications management determines the best channels (mobile application, voice, SMS, web, email) and times to contact customers, then orchestrates the process. Real-time updates to/ from other streaming analytics and customer-facing systems ensure coordination. In collections, FICO clients using this technology have achieved 30–45% reductions in closing balances; in fraud management, 35–70% reductions in losses.

**Conclusion**

There’s no doubt we’ll continue to see remarkable advances in our ability to capture, store, access and analyze Big Data. But its value in advancing the top priority for many businesses today—customer centricity—depends on our ability to systematically use it to learn more and more about our customers.

Accomplishing this depends on judicious use of the powerful automated technologies that Big Data computing infrastructures are now making available and practical for most businesses. It also depends on human expertise, an absolute essential to ensuring that these technology investments pay off in useful insights about customers.

In fact, the importance of such expertise will only increase as we move forward. At some point, Big Data technologies will be so efficient that they’ll become a common denominator. The differential will be in the human experience, talent and ingenuity that lead us to interpret and apply data analytics in new and more competitive ways.

Dr. Andrew Jennings is a Senior Vice President and Chief Analytics Officer at FICO. Since joining FICO in 1994, he has worked with leading institutions worldwide to raise analytic performance. Previously he held senior analytics positions at Barclays and Abbey. Dr. Jennings has a Ph.D. in economics from the University of Nottingham. He regularly blogs on the Banking Analytics Blog and FICO Labs Blog.