# The data dilemma

Unlocking customer insights with machine learning

Many financial institutions (FIs) see their customer data as one of their most valuable assets. Unlocking insights from that data helps FIs understand, anticipate, and offer account holders the products and services they truly need. A major trend to unlocking customer insights is using machine learning to surface the behavioral intelligence buried in the large amount of account holder transactional data captured each and every day. In this paper, learn how a group of talented, enthusiastic analysts with an open approach to data can yield some very interesting and extremely valuable and actionable results. This approach, championed by Q2's CTO Adam Blue, has led to a new platform, Q2 SMART, which provides powerful behavioral analytics for FIs, enabling growth while providing account holders with real value.

By Adam Blue, Chief Technology Officer, Q2

### Introduction

Some stories, however clichéd, are just too instructive to be abandoned. My favorite of these stale chestnuts is the story of the blind men and an elephant:

The story of the blind men and an elephant originated in the Indian subcontinent from where it has widely diffused. It is a story of a group of blind men (or men in the dark) who touch an elephant to learn what it is like. Each one feels a different part, but only one part, such as the side or the tusk. They then compare notes and learn that they are in complete disagreement. In this age-old parable, the treachery of perception is laid bare as each man has a completely legitimate but isolated experience with the creature, and thus cannot agree on what it is. This is not an uncommon occurrence in organizations-even among the sighted—where the metaphorical blindness of inherent positional bias and organizational inertia impact what each of us "sees." Where a marketer might see opportunity, a salesperson might see an exhausted market. A service rep sees ways to find and solve problems, and an IT person sees only more work on his or her plate. Turning to "objective analysis" of data seems like it might be a way to resolve these differences across stakeholders, but all too often the collected data only ends up confirming the pre-existing bias each person holds.

As Q2 performed research about the ways in which end users utilized online and mobile banking and banking services in general, we worked to eliminate this pre-existing bias from our efforts. By looking at the underlying data and allowing it to speak for itself through models constructed via statistical analysis and supervised machine learning, we worked to separate the data from the assumptions and "conventional wisdom" that can obscure real insight. The results were instructive, and this paper will examine some of the techniques

as well as the infrastructure and compute architecture used to build an ecosystem of analytic capabilities to support ongoing research.

Source: In Wikipedia, Retrieved September 2016, from https://en.wikipedia.org/wiki/ Blind\_men\_and\_an\_elephant

### Yard sale data

#### Finding treasures

Many organizations have large amounts of data as a result of operating and tracking business processes that can have tremendous value. Unfortunately, this data resembles a neighborhood yard sale: The data is scattered and no one knows what it is or from where it came.

This problem of fitting the data together across customers, accounts, periods of time, and business processes is a significant one. When Q2 set out to understand what end users were doing in the digital channel and in their everyday lives, we first began with a small set of well-understood data (in our case, commercial and retail inter-institution payments), and then worked our way outward. At each step when we added more data to the overall set of models, we were careful to ensure we could link it directly to the end user and map this data into the appropriate time periods of observed behavior. Since Q2 wanted to eliminate the uncertainty that comes from data quality issues associated with many "data warehouse"-style projects, we excluded from our model data that could not be associated definitively with an end user.

Data can resemble a neighborhood yard sale: The data is scattered and no one knows what it is or from where it came.

### Making the highly technical consumable

### Data abstraction can help

Once Q2 arrived at a pipeline of data from our product platform, we approached the challenge of how to work with it. Most importantly, we asked the question: "How can non-IT staff work with this data to uncover insights and learn about their customers?" There can be a real barrier in typical organizations between the team members who think about business and process issues, and the team members who manage and analyze data. Building a method for business-focused staff to consume the collected data in a way that did not require development or IT skills was paramount.

As our data scientists worked with the data and put models in place to structure how we think about customer behavior, it became apparent we were creating new metadata that captured the value of the underlying observations at a high level. This is a very common technique in data analysis, including for instance using the percentage measure of favorable or unfavorable responses in a survey or poll rather than the raw numbers. Another great example would be characterizing an athlete's performance in terms of a statistical measure of achievement, such as the batting average for a hitter in baseball. Using these abstracted measures to aggregate a large amount of data into a compact concept provides a means for discussion and comparison. Q2 calls these new measures "traits," and we compute them from the large set of underlying data that we have available.

How can **non-IT staff** work with this data to uncover insights and **learn about their customers?** 

We can then use these traits to investigate or target groups of customers or end users, looking for correlations and trends at a much higher level of information than we could if we were bound directly to the low-level data. Some traits are simple descriptions of the level of observable behavior. Measures of feature and channel or device usage such as bill pay or mobile banking transactions are examples of simple traits. We also model the likelihood that a customer might prefer a particular banking product as his or her next best product as a trait. (We'll talk about how we arrive at this measure in the next section.) Although we might use very different techniques or widely varying data, this abstraction model of mapping everything we know to metrics that answer the question, "How often or how frequently does this customer exhibit this behavior?", allows Q2 to rationalize enormous amounts of data into a manageable set of elements.

## Collaborative filtering for next best product

# Creating a Netflix-like experience for recommending financial services solutions

One of our first research projects was to determine if Q2 could identify customers who would be more likely to benefit from and be interested in a particular banking product based on an overall analysis of the products customers held across the base. We applied a technique known generally as collaborative filtering to make predictions about end user preferences using the collected information across the set of users.

It is useful in this case to think of the set of products a customer has with a bank or credit union as an expression of his or her preferences about the set of available products. For instance, Netflix has done extensive work in using preferences for film and TV content to make content recommendations their end users would enjoy. In considering the underlying drivers that impact a consumer's choice of entertainment content or bank products, it is reasonable to hypothesize that similar consumers would make similar choices. Given that there is a set of data that characterizes their preferences as expressed so far—such as movies already viewed or products already owned—the collaborative filtering model identified the best choices for customers based on the choices similar customers have already made.

In our experimental exercise, Q2 identified which customers would be the most likely to prefer a new account of a given type, in this case an auto loan. Then we measured whether or not the customers in the selected group opened new auto loans with the financial institution in the exercise. The identified users from the collaborative filtering model were three times as likely to open the new product as a user from the larger, more general group, indicating the algorithm for selection has strong predictive power. Given that this technique can help us focus on which customers to send the right marketing messages, content, or offers, it allows for a critical tailoring of the approach to driving product adoption. In this way fewer, more targeted communications can lead to increased conversions.



The identified users from the collaborative filtering model were **three times as likely** to open the new product as a user from the larger, more general group.

### A model of account balances

# Helping account holders map their financial journey

Customers care about the balances in their accounts, both as a means of understanding their current financial positions and, over time, their financial journeys. Traditional approaches to alerts ("Tell me when my balance is below \$500") can be useful, but they require an end user to establish a rule for this behavior and then choose a threshold. This threshold is also fixed, so for a given user having only \$500 left just after getting a paycheck could be cause for panic, while having \$500 left just before getting a paycheck is cause for celebration.

One of Q2's research goals was to create and operate an algorithm that can model the behavior over time, providing an ability to handle the scenario above as well as understand balances more generally. Research into the transaction patterns that occur over time for customer accounts was conducted using a variety of techniques (most notably Fourier and wavelet analysis) that would allow analysis of the transaction pattern without presupposing what the patterns represented.



Using these mathematical tools, it was interesting to see that these methods are very good at detecting recurring patterns in customer behavior that allow for modeling of the balance over time. For your customer, this means it should be possible to **send an alert** when their **balance** is unexpectedly **low**.

As a result of the modeling, it is possible to identify upward and downward trends in spending and income as well as to determine when their balance that results from transactions is outside of the expected range for the position in the recurring period of activity. For your customer, this means it should be possible to send an alert when the balance is unexpectedly low for a given point in the financial cycle. This approach should prove to be a much better experience since it does not require the end user to configure the alert, it allows for the algorithm to dynamically adjust to the point in the income/spending cycle that the balance is checked and the alert itself is tailored to the specific behavior of the customer over time. Our research in this area has also resulted in a better understanding of the broad classes of behavior into which customers fall. Clustering analysis of the patterns that are exhibited by analyzing the deposit and spend patterns in operating checking accounts shows several distinct clusters of customer behavior that correlate well with other observable characteristics, such as income or family status. These items can then in turn be mapped to traits as described above that allow for tailoring communication, offers, or surveys to customers that exhibit specific behavioral patterns.

# The real answers lie in asking the right questions to drive new technology



Q2 focused on the questions we wanted to ask and let that drive the selection and construction of any new tools. There has been much fevered prose written in recent years about "big data" and the tools used to work with it (Hadoop and Elasticsearch come to mind). Invariably, much of the attention has been on solving technology scale problems associated with working with very large amounts of data. All of these tools, however, were built to solve specific problems the authors of those software systems were trying to solve. In our environment, Q2 focused on the questions we wanted to ask and let that drive the selection and construction of any new tools.

At the core of the approach for us was the implementation of an Apache Cassandra cluster, which gives us a scalable, fault-tolerant, high-performance NoSQL data store. We then constructed a pipeline to extract, transform, and load data from the operational data stores that drive the digital banking platform. This pipeline is where we were able to add a lot of value between the existing platform and the Cassandra instance, particularly because we have a good understanding of the semantics of the source data. This pipeline, which we call "wheelbarrow," gives our data scientists a low-cost way to gather and rationalize data from across the organization and transform it as it flows into the primary data store. Then, it is available for extensive analysis without any impact to the operational systems.

Last, Q2 made a key decision in planning for this infrastructure to create an API surface for the other parts of our development team to consume this data. In this way, the data science group can work independently from the rest of the team without worrying about breaking production processes or impacting performance. Similarly, the compute burden of mapping the data is done offline outside of the core product platform.

### From iterations to evolution

Now that we have arrived at several models of behavior and we have an abstraction of analytics products in the form of traits to present to others, Q2 is focused on continually learning more about end users and tuning the analytics we have in place. As we evaluate the effectiveness of the models for identifying the next best product for instance, observing the outcomes of the campaigns can lead to re-examination of the models and identification of additional data elements to add to the modeling.

Over time, Q2 expects to derive a significant amount of improvement in modeling and prediction from watching our bank and credit union customers use the platforms we provide, then questioning assumptions, rethinking models, and trying alternate techniques. Besides Q2's investments in the technology back end to allow for rapid iteration, partnering with our FIs to observe what is working and not working is crucial. We used this same methodology in the initial construction and subsequent improvement of Risk and Fraud Analytics, our real-time fraud prevention module for digital payments in the online and mobile channels.

Many organizations in the financial services space claim their data is one of the most valuable parts of their businesses. However, the investments made in actually performing primary research frequently do not match that claim. Conventional wisdom and the "elephant parable" that opens this paper are prevalent even in progressive, high-functioning organizations. Our experience has found that a few talented, enthusiastic analysts with an open approach to the data can yield some very interesting and extremely valuable results.



Over time, Q2 expects to derive a significant amount of improvement in modeling and prediction from watching our bank and credit union customers use the platforms we provide.



Adam Blue Chief Technology Officer, Q2

#### About the author

Adam Blue is Q2's Chief Technology Officer. His passionate belief that technology can and will change the lives of people, businesses, and financial institutions everywhere is what sets Adam apart from other CTOs. An established player in global IT, he brings more than 15 years of technology and software management expertise to Q2. Adam holds a Bachelor of Science in Economics from Indiana University. He has also completed graduate work in Computational Economics at the University of Texas.

#### About Q2

Q2 is a leading provider of digital banking solutions, headquartered in Austin, Texas. Driven by a mission to build stronger communities by strengthening their financial institutions, we provide clients with the industry's most comprehensive and adaptable digital banking platform as well as access to open technology, development tools, and actionable, data-driven insights to boost account holder experiences and retention, while creating more growth opportunities.

To learn more about Q2, visit <u>www.q2ebanking.com</u>.