Bank(ing) on Data Science

By embracing data science tools and technologies, banks can more effectively inform strategic decision-making, reducing uncertainty and eliminating analysis-paralysis.

Executive Summary

Amid the ever-present big data buzz, some large global banks have mastered the art of using data science to improve customer engagement, revamp products and optimize marketing outreach, risk management, pricing and ongoing cost reductions. Meanwhile, others are still trying to make sense of where these emerging technologies and techniques fit in. At some point, banks of all sizes, shapes and forms need to incorporate data science into their operating models.

The future of banking will be determined by how well banks use technology to maximize their accumulated wealth of transactional and interactional data to better understand hidden patterns of customer behavior. Using these insights, they can make necessary service improvements and customize existing offerings to properly align the right products with the right customers.

To successfully implement data science, banks need to start small and adopt a structured approach, based on a strategic roadmap. Banks that can analyze the data they collect and utilize it for strategic decision-making will maximize their competitive advantage; those that cannot will place their profitability, if not their survival, at risk.

By understanding data and applying insights gleaned from customers, partners and employees, banks can more effectively compete on code and gain incredible competitive advantage. Companies such as Google, Pandora, Netflix, Amazon — and many others — are winning decisively in their markets because of their refined ability to mine insight from the digital information surrounding people, organizations and devices, or what we call a Code Halo™. When properly harnessed, Code Halos contain a treasure trove of business value.1

This white paper details the growing importance of Code Halos in data science and analytics initiatives. Importantly, it highlights potential areas of fit, ways to overcome challenges and a recommended implementation strategy for key data science initiatives.

Banking’s Evolving World

In the aftermath of last decade’s global financial meltdown, the banking industry is undergoing a radical transformation due to rapidly changing consumer behaviors and expectations; more stringent regulatory guidelines; and a highly competitive environment with a proliferation of new channels (mobile banking and social media) and competitors (nonbanks, such as Paypal and Google Wallet).

This ongoing transformation, while difficult, is also opening doors to new opportunities. Banks must ensure that they can cost-effectively acquire new customers while retaining existing ones. And...
to expand their reach and profitability, they must also tighten their focus on the expanding digital world. Analytics, big data and data science can unlock a world of new possibilities. With proper use of data science, banks can better understand prospect/customer relationships by exploring ever-changing transactional and interactional behaviors. New digital marketing technologies, such as Web sites, e-mail, mobile apps and social networks, are helping banks better target their customers and improve engagement. Moreover, advanced segmentation strategies are helping them boost their marketing effectiveness by identifying niches based on consumer behavior.

**Growing Importance of Data Science**

The goal of data science is to extract hidden insights and knowledge from data. In our view, the key word is “science,” since, done properly, data science requires a systematic study of observation, backed by proven scientific techniques. Data science builds on elements, techniques and theories from many fields, including signal processing, mathematics, probability models, machine learning, computer programming, statistics, data engineering, pattern recognition and learning, visualization, uncertainty modeling, data warehousing and high-performance computing. The exponential growth of data, particularly unstructured data, makes big data an important aspect of data science. Every day, 2.5 exabytes of data are created; just one exabyte is equal to 50,000 years’ worth of DVD-quality video.

For years, financial institutions have leveraged customer insights gleaned from systems of record to manage risk and fraud, as well as to improve product development, marketing and customer communications. Today, new and enhanced technologies, coupled with the availability of a vast pool of structured and unstructured data, allows for real-time, multichannel decision-making processes that can save money and increase revenues.

Many banks are just beginning to consolidate and utilize the internal data elements at their disposal, such as debit and credit transactions, purchase histories, channel usage, communication preferences, loyalty behavior, etc. And when it comes to big data, banks have collected large amounts of information from a variety of sources, such as transaction details and spending behaviors. The addition of newer sources, including Web server logs, Internet clickstreams, social media activity and mobile-phone call details, has opened the floodgates on the data sets that can be mined for meaning.

However, this is easier said than done, as these data sets come in a variety of structured, semi-structured and unstructured formats, and arrive at an ever-increasing velocity and complexity. Analyzing this data is now mission-critical, since it can provide more timely and precise insights to guide business planning and decision-making. With so much transparent content generated daily through social media, data science can help banks deliver a consistent and integrated customer experience.

To use this data for business advantage, banks must set up data analysis teams to collect, sift and apply meaning from this data to advance business goals. According to Gartner, big data in the banking industry has the highest level of opportunity because of the high volume and velocity of data in play. Moreover, 78% of CFOs have labeled BI and analytics as the top technology initiative for their departments – beating out even financial management applications.

**Key Inputs for Data Science**

As noted earlier, data can be broadly categorized as structured and unstructured. At a broad level, structured data comprises transactional data, which includes customer buying/spending habits, and unstructured data can be obtained from various social media sites, such as Facebook and Twitter. Precise analysis of social data is of great importance because it provides valuable insight into individual customers’ likes, dislikes, preferences, etc.

Analysis of both structured and unstructured data can help banks better target the right product to the right customer at the right time. For example, by correlating the social activities (unstructured data) of a customer with a spending pattern (structured data), banks can customize and optimize the timing of their product offerings.

For even more precise targeting, organizations can add new third-party data sources, compiled from a variety of sources, such as public repositories, mobile devices and cars. As such, data science involves three aspects of data: velocity, volume and diversity (see Figure 1).
Data Science: Usage Areas

Many business areas can benefit from data science (see Figure 2). To properly ascertain how customers prefer to be served, banks can apply such data science techniques as hypothesis testing, crowdsourcing, data fusion and integration, machine learning, natural language processing, signal processing, simulation, time series analysis and visualization. Using the insights gleaned from these approaches, marketers can derive the right marketing strategy through a mix of marketing messages and offers that resonate with individual customers and segments.

For example, using a mobile app, banks can analyze individual consumer behaviors and spending activities and combine that data with credit bureau information. When analyzed, the resulting insights can lead to better targeted messaging around a potential offer, such as a pre-approved home loan to a customer who is qualified based on analysis of the data contained in his transactional files and interactions on social media.

The vast amounts of online data have much to offer banks seeking consumer insights. For instance, by combining information from travel Web sites and spending patterns gleaned from internal databases, banks can optimize their product mix and offers. Analysis of transactional behavior like recency, frequency and monetary value can be sliced and diced to derive customer profiles that can improve the effectiveness and efficiency of targeted marketing efforts. An example is an Australian bank that is working with a retailer to better understand where the retailers’ customers live, when and where they shop, and how much they spend. This information is then used to refine the retailer’s branch location/relocation strategy.

Another example is a bank that uses point of sale data to determine whether a customer frequently enters a certain area for shopping or lunch and then use this information to deliver offers that are highly personalized even to the type of food the customer prefers, increasing the probability that the offer would be accepted. Adding device-specific capabilities, the offer could be delivered by SMS at the most logical time for decision-making.
As an early warning system, data science solutions can help banks quickly identify potentially fraudulent behavior before the fraud becomes material. For example, individual cardholders are creatures of habit. Cardholders have “favorites” or recurrences over a wide variety of objects in their transaction streams. These objects might include favorite ATMs that are close to work or home or gas stations along a daily commute, as well as preferred grocery stores and online sites for shopping.

An analytics technique that could be used to improve fraud management is to identify cardholder favorites, in order to distinguish between “in-pattern,” or normal, customer spending and “out-of-pattern” suspicious transaction activity. This enables faster fraud detection at much lower false positive rates (declines on legitimate transactions).

Text analytics of unstructured data can help banks identify patterns of information that indicate the likelihood of fraud. Text mining of insurance claim descriptions (written and recorded) provided by bogus claimants uncovered some very interesting facts. It turns out that certain phraseologies (the use of “ed” rather than “ing” on the end of verbs, for instance), are extremely indicative of fraudulent claims. This is due to the different ways in which people relay stories they actually experienced vs. those they concocted; for instance “I was walking” is indicative of someone recounting an actual experience whereas “I walked” often turns out to be indicative of someone describing a fictitious event.

Unstructured data, such as social media comments, can help banks gain insight into what customers like and don’t like about various brands, products and service and also gather feedback about their own products and services. By closely tracking customer comments, banks can quickly identify issues and take action to improve the customer experience. The instant feedback of social media also enables banks to capitalize on opportunities to proactively counteract negative perceptions, exceeding customer expectations and driving loyalty. Banks can also use social media data to target customers with offers or services aligned with recent life events (e.g., graduation, marriage, new job).

Data science can help banks recognize behavior patterns, providing a complete view of individual customers and segments. For example, when a customer enters a bank, customer representatives can be better equipped to offer the right products and provide a quicker resolution to customer queries by analyzing their Code Halos. Data science can also be used by banks to analyze the average cost for each channel (e.g., call center, branch banking, etc.) and design strategies to migrate customers to low-cost channels.

Analytics techniques can also play a significant role in the early warning, detection and monitoring of fraud. These techniques allow organizations to extract, analyze, interpret and transform business data to help detect potential instances of fraud and implement effective fraud monitoring programs (see sidebar).

Advanced data science techniques could enable institutions to improve underwriting decisions and increase revenues while reducing risk costs. These techniques can be fruitful across all asset classes, all types of credit risk models and the entire credit life cycle, including profit maximization and portfolio management.

For debt collections and recoveries, analytics is a critical part of the process, as it can enable organizations to create an accurate picture of the customer’s propensity and ability to pay and, therefore, the amount likely to be recovered. This behavioral scoring is used to segment customers and prioritize collections activities to maximize recoveries and reduce collections costs.

**Overcoming Challenges**

What follows are the common obstacles banks encounter when attempting to implement an effective data science strategy.
Data Volume
Over the last decade, banks have accumulated huge volumes of data, especially following the introduction of smartphones, tablets and now wearables that enable multi-channel access; however, many still suffer from a scarcity of insight. Managing enormous data sets, as well as analyzing and correlating structured, semi-structured and unstructured formats, makes the data science job increasingly complex.

Distinguishing “signal” (meaningful insight) from “noise” (massive amounts of unmanaged data) remains a fundamental challenge and a significant opportunity. There are various data cleansing techniques, such as clustering, outlier detection, etc., that can help organizations find correlations within data sets.

Budget Constraints
Banks must be willing to invest significantly in people, infrastructure and platforms to effectively analyze and make strategic decisions from big data. Beyond these investments, such initiatives also need to strategically align with the bank’s overall vision and business mission. Such initiatives require qualitative and quantitative scrutiny in order to prioritize the projects with the highest payback. Priorities can be determined by strategic and tactical benefits, cost, duration, people and technology availability.

Privacy Concerns
Gaining permission to use and process data from mobile and social media is a huge challenge. Numerous concerns have been raised over identity theft, privacy and social media stalking, among other issues. Within the bank, it is also important to ensure that the right people across the organization (i.e., bank decision-makers) can access the right data, at the right time.

Organizations must also decide who owns the data before a data science project is implemented, so that accountability and workflow can be properly set and followed.

Skilled Talent
There is a huge demand for data scientists, and the pool of available talent is insufficient to meet the needs of every organization. Finding highly skilled data scientists is not easy; they do not simply report on data but look at it from many angles, running complex queries to find correlations and patterns. They also need to communicate their findings and recommendations to top leadership. Some of the top skills required for data scientists include analytics know-how, statistical acumen, domain expertise data mining and the ability to clearly and effectively communicate.

Looking Forward
Today’s knowledge economy provides businesses of all kinds with access to big data that’s growing exponentially in volume, variety, velocity and complexity. With more data coming from more sources faster than ever, the questions will only continue to unfold. Some examples:

- What is your organization’s data science strategy?
- How is your enterprise combining new and existing data sources to make better decisions?
- How could new data sources, including social, sensors, location and video, help improve your organization’s business performance?
- Will your organization take advantage of big data or remain paralyzed through endless analysis?

A savvy, experienced team of data science consultants can help organizations create a roadmap that results in a meaningful, business-aligned approach to data science. Experts can help implement data science technologies, manage big data, accurately predict customer demand and make better decisions faster than ever before.

The best approach is to start small rather than setting off a big bang. The mantra for successful data science projects depends on the organization’s business objectives, but one constant is focus and agility. For example, if the business need is to define customer segments to drive pricing-elasticity models, the IT organization should first discover which customer data needs to be gathered before building an enterprise data warehouse and an enterprise analytics platform. Experts can develop an initial proof of concept by analyzing the internal, external, structured and unstructured data and conclude with meaningful, business-aligned recommendations.

A blueprint can help guide the organization to develop and implement data science solutions in ways that deliver business value. From there, an implementation strategy followed by a detailed plan can be built (see Figure 3, next page).
Data Science Implementation Plan

Figure 3

Footnotes


About the Authors
Shantanu Dubey is a Consultant within Cognizant Business Consulting’s Banking and Financial Services Practice. He has over seven years of business and IT consulting experience in implementation of BASEL regulatory reporting, business intelligence and core banking solutions. Shantanu also has experience working with leading banks on product development, business process optimization, business requirements management and gap analysis across various geographic locations. He holds a bachelor’s degree in information and technology engineering from RGPV Bhopal, and a post-graduate diploma in management from I2it Pune. Shantanu can be reached at Shantanu.dubey@cognizant.com.

Siddhartha Nainwani is a Consultant within Cognizant Business Consulting’s Banking and Financial Services Practice. He has over seven years of business and IT consulting experience, working with leading banks in business process management, business analysis and test management across various geographic locations. Siddhartha holds a bachelor’s degree in engineering in information technology from Shivaji University, Maharashtra, and a master’s degree in management from ICFAI Business School, Mumbai. He can be reached at Siddhartha.Nainwani@cognizant.com.
About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting, and business process outsourcing services, dedicated to helping the world’s leading companies build stronger businesses. Headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfaction, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With over 75 development and delivery centers worldwide and approximately 178,600 employees as of March 31, 2014, Cognizant is a member of the NASDAQ-100, the S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world. Visit us online at www.cognizant.com or follow us on Twitter: Cognizant.

© Copyright 2014, Cognizant. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the express written permission from Cognizant. The information contained herein is subject to change without notice. All other trademarks mentioned herein are the property of their respective owners.