You're Doing It Wrong: How to Get Data Science Right

A blueprint for extending analytics into the business

In the critical "last mile" of connecting data insight to action that drives bottom-line results, too many organizations are doing it wrong.

As organizations continue to seek ways to use the huge volumes of data they are capturing to improve efficiencies, boost customer service, and gain a competitive edge, data science and advanced analytics continue to grow in importance. There's no question that big data is big business: IDC forecasts big data and business analytics software will reach $189.1 billion in 2019 with double-digit annual growth through 2022.[1]

But are all of these investments good business? According to a recent global McKinsey executive survey on analytics activities in a range of industries, "86 percent of executives say their organizations have been, at best, only somewhat effective in meeting the primary objective of their data and analytics programs" while "one-quarter say they've been ineffective."

Forrester has reported that "while 74 percent of enterprise architects aspire to be data-driven, only 29 percent say their firms are good at translating the resulting analytics into measurable business outcomes." Part of the problem is that in this critical "last mile," connecting data insight to actions that drive bottom-line business results, too many organizations are doing it wrong.

---

1 https://www.businesswire.com/news/home/20190404005662/en/IDC-Forecasts-Revenues-Big-Data-Business-Analytics

Healthcare Case Study: Evicore

Define

Health solutions provider wanted to better leverage incoming data to speed the benefits approval process.

Transform

eviCore extracts structured and unstructured data, including patient records, provider notes, insurance codes, and more.

Model

Predictive model created in TIBCO Data Science crunches data to score claims based on the likelihood of approval or denial.

Deploy

Scores are integrated into the approval process workflows that staff uses to evaluate claims.

Act

Based on score, claims are either automatically approved or sent directly to MDs for further evaluation, saving process cycle time and resources.

Evaluate

500% process improvement achieved, the next swing of pendulum will extend use of the model to other claims categories and business units.

To get greater value from analytics, organizations need to view data science not as a project, but as a process that includes a range of stakeholders and a repeatable set of steps, including feedback loops that engender continuous improvement. This whitepaper looks at some of the key analytics mistakes businesses are making, and outlines a better approach, a blueprint that will help ensure more streamlined, and ultimately more successful, analytic efforts.

Where Do Organizations Go Wrong?

Even before "big data" became a common buzzword, organizations were investing in technologies to extract value from their data. But all too often, analytics projects fail to deliver tangible business results. These failures to successfully transform data into action spring from several common mistakes:

No business objective

Analytics, in and of itself, is not an objective. Rather, it is an enabler that empowers the organization to achieve business objectives—personalizing the financial advice that wealth managers give their clients, speeding an insurance claim approval process, improving the performance of digital marketing campaigns. Yet many companies lose their way because they focus first on technical considerations like databases, software, and infrastructure before getting input and sponsorship from the business.

Business owners are the best people to define the problems that analytics needs solve, identify the end users that solutions must be delivered to, and suggest data sources that can be tapped for insight. To avoid costly mistakes, these stakeholders must be part of the analytics strategy from day one, because no matter how seamless or technologically advanced, deployments with vague or poorly defined objectives will fail to drive business results.

The "magical thinking" problem

Issues can also arise when the business side of the organization views the analytics practice as a "black box," in other words, a place where data scientists work their magic crunching numbers. Without a basic understanding of what types of analyses can be done on different types of data and why, business users will never know the true value of the information they have at their disposal or how they can use it to improve their work. The organizations that capitalize on their data most successfully (think of companies like Amazon) are those that make an effort to demystify data science and treat analytics as a core practice, just like sales or marketing.

Cult of the algorithm

There's a lot of talk about algorithms in the analytics space, with vendors making claims about the sophistication and performance of their models relative to competitors. But the fact is that algorithmic complexity can easily come at the expense of operationalizing analytics in a timely fashion. For organizations looking to quickly transform data into action, the saying "don't let the perfect be the enemy of the good" is an apt one.

Oftentimes the most traditional machine learning techniques like regression and clustering are enough to solve a business problem, while highly complex algorithms only add to analytic cycle time and expense, without adding value. This is not to say that there isn't a place for more sophisticated analytic models, but sometimes organizations waste a lot of time and money building complex models when a simpler approach would do.

No process

Stating an objective is one thing, having a solid plan in place to reach that objective is another. Because analytics is often viewed as an IT or data science project rather than as a process, organizations can easily get bogged down trying to implement a solution whose purpose hasn't been clearly defined, that doesn't have buy-in from the right people, or that is ultimately too complicated for end users to benefit from.

A business-focused analytic strategy requires that organizations follow a methodology designed to ensure that specific milestones and objectives are reached. Without steps to capture business requirements, ensure data quality, operationalize insights, and measure results, analytic success will be spotty at best. And even if taking a project-based approach does yield some insight, without a repeatable process in place, it will be difficult to scale the analytics model to other areas of the business.

The Pendulum: A Blueprint for Boosting the Value of Analytics

It's clear that organizations need to be more thoughtful and systematic about how they incorporate data science into their business. Analytics, to be of any value, must be operationalized at the point of action, whether that means prompting a sales call, automating a purchasing request, or providing insight that drives better real-time decisions. But putting data into action consistently doesn't happen by accident. Organizations need a blueprint that they can use time and again to set themselves up for success, and to avoid pitfalls that result in investments that fall short of expectations.

Financial Services Case Study

Define

Global financial services firm wanted to tap data to customize wealth management advice.

Transform

The firm created an "insights engine" to evaluate internal and external data spanning millions of clients, market events, and investment positions.

Model

TIBCO Data Science software enabled the firm to build models that score investment opportunities based on relevancy to specific clients.

Deploy

Data insights are made directly available to wealth managers via the applications they use daily.

Act

Wealth managers can give their clients more personalized advice based on science rather than intuition and evaluate new opportunities in minutes.

Evaluate

Scoring models are constantly evaluated, and feedback loops allow wealth managers to weigh in on what insights are driving value.

TIBCO is focused on delivering a data science deployment process that resembles a pendulum. This process starts and ends with the business, swinging forward and back through a series of steps designed to ensure that people and technology are organized to deliver the greatest possible value from advanced analytics. The forward swing of the pendulum focuses on driving data to the point of action as quickly as possible:
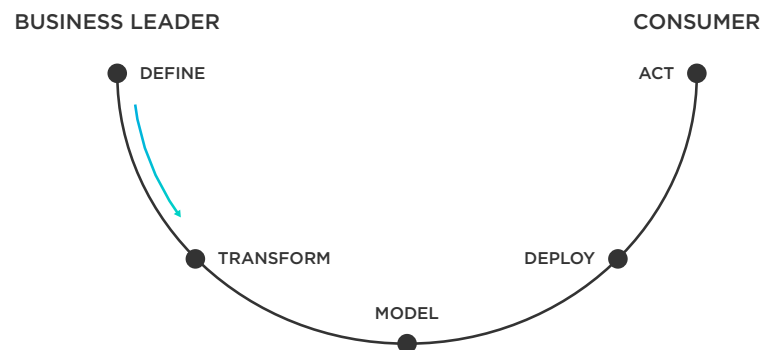
Define

In this critical first step, the business owner must define the problem that needs to be solved, specifically focusing on the behaviors (sales calls, for example) that will be impacted by analytic insights. Key questions to ask during this step include: What data sources might help solve this problem? Who are the business (or customer) end-users? What applications and systems do they use to do their jobs?

Transform

Once the objective is clearly defined, data engineers need to connect to the data to extract the information the business deems relevant. A key focus here is data quality issues such as siloed information, inconsistent taxonomies, and integration that must be addressed before moving on to analytics.

Model

During this step, data scientists build models that address the problem statement. Here, it is important to remember the "cult of the algorithm" pitfall. Operationalization takes precedence over model sophistication, so it doesn't make sense to spend too many cycles developing the "perfect" model before testing whether or not it delivers value.

Deploy

This step focuses on the critical "last mile," connecting data insight to actual business outcomes and end user actions. Operations and application engineers create APIs and app integrations to embed scored model output at the point of action.

Ad-Tech Case Study: Havas Media

Define

Global media agency wanted to empower account teams and clients to easily access campaign data and better understand performance advice.

Transform

600 TB data warehouse centralizes huge volumes of media and campaign data.

Model

TIBCO Data Science software replaces labor and time-intensive analytic process with drag-and-drop workflows that allow analysts to quickly create and deploy models.

Deploy

Account teams and clients can now access data, launch, and visualize analytic queries in a self-service environment supported by TIBCO Data Science software.

Act

Data insights are available at the point of decision, empowering campaign adjustments and refinement to be done on-the-fly.
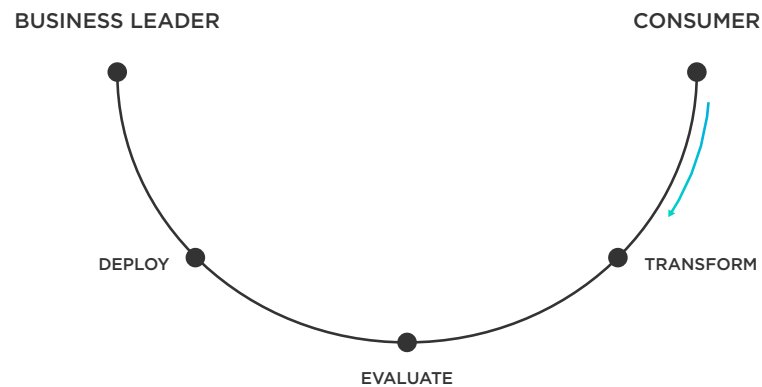
Evaluate

Successful analytic models become part of a library, and can be reused by multiple account teams across the organization.

Act

The business now behaves differently, because internal and/or external end users are empowered to make better and faster decisions. According to the McKinsey global survey on analytics activities, organizations having the greatest success with their initiatives were "more diligent than others when it comes to measuring results." The survey showed that 54 percent of high performers tracked the impact of their analytics activities on top-line revenues, while only 19 percent of respondents from low performing organizations did the same. This is why collecting data on the business impact of the analytic model is a critical part of the pendulum process.

The backswing of the pendulum focuses on capturing feedback for continuous improvement:

Measure

Measure model accuracy end user compliance with model suggestions, and business outcomes.

Evaluate

Determine whether the change in behavior driven by the model has led to business value.

Report

Deliver results back to the business owner that kicked off the project.

Once the feedback swing is complete, organizations can choose to do one of three things based on the observed outcome of the analytics deployment:

No value or negative value

Stop the pendulum and take the model offline.

Positive value, little room for improvement

Stop the pendulum and leave the model operationalized.

Positive value, substantial room for improvement

Give the pendulum a hard swing and start the cycle again.

A key thing to remember is that failure is an option because the pendulum process is designed to constantly capture feedback on what's working and what's not. In this way, analytic models can be continuously tested and refined. Like agile software development, the focus is on rapid deployment followed by additional rounds of incremental improvement. This aligns with the TIBCO philosophy: analytics is not an objective addressed by a single project, but rather an ongoing organizational process that may serve a number of business objectives over time.

TIBCO Data Science: A Platform for Business-driven Analytics

In a report on data and analytics, Forrester emphasizes the importance of shifting from "data-driven priorities to insight-driven execution," a sentiment that mirrors TIBCO's focus on operationalizing analytics at the point of action. The report states: "… the most successful firms will push beyond the limits of agile BI and big data by aligning business and technology leaders toward a unified goal. They will connect data, insight, and action and continuously learn what works—and what doesn't."

TIBCO Data Science software was built with this call to action in mind. The TIBCO Data Science Platform helps organizations solve business challenges with data by providing a repeatable process for operationalizing advanced analytics. Unlike other solutions that focus primarily on algorithms, TIBCO Data Science software brings machine learning, data, and people together in a centralized environment, enabling both business and IT stakeholders to take part in transforming insight into action.

The TIBCO Data Science Platform empowers organizations to move through all of the steps outlined in the pendulum blueprint for driving analytics to the business. It begins by helping business owners define the problem they'd like to solve. Through distributed analytics services, the TIBCO Data Science solution then provides the data transformation and modeling capabilities that enable advanced analytics to be run quickly and efficiently on large volumes of data at scale. Finally, TIBCO Data Science software helps operationalize and connect machine learning assets to action by extending outputs directly into existing business user workflows and end point applications.

Several key features come into play to make this possible:

Collaborative platform

TIBCO Data Science software provides a collaborative environment that engages both business and IT stakeholders in the data science process. The platform is designed to help business owners define their analytic objectives in terms that can be translated into machine learning models, and to organize the processes and people (ranging from data scientists, application developers, business end users, and others) that need to come together to deploy successful solutions.

Simple-to-deploy analytics

The TIBCO Data Science analytic engine provides standard, proven algorithms for regression, classification, clustering, and more in a simple-to-deploy, drag and drop environment. Organizations can build and operationalize analytic models quickly and with little IT involvement. The TIBCO Data Science solution also offers extension capabilities that enable customization of models if and when more complex analysis is required.

Last mile integration

The TIBCO Data Science SDK allows organizations to push analytic outputs directly into the applications that business users rely on daily to do their work. With the SDK, analytics can be integrated into existing workflows and applications in forms that are immediately recognizable, such as a text or email alert, so that business users can easily benefit from machine learning without having to be a data scientist or trained on an unfamiliar system.

The Democratization of Data Science

Data may be everywhere, but if organizations continue to view data science as something only PhDs can understand, the true potential of advanced analytics will continue to remain untapped. The companies that are most successful with data science focus on creating a "culture of analytics" that transforms raw data into actions that have direct impact on customer service, process efficiencies, competitive advantage, and more.

By delivering a process that starts and ends with the business, TIBCO Data Science software ensures that machine learning efforts don't end up locked in a black box as isolated research projects with no impact on business behavior and outcomes. Getting data science right demands an approach that engages both IT and business stakeholders and that focuses on tangible outcomes and continuous improvement.