

Mining Your Own Business

A Primer for Executives on Understanding and
Employing Data Mining and Predictive Analytics

CHAPTER 3: Leading a Data Analytics Initiative

JEFF DEAL

GERHARD PILCHER

Foreword by Eric Siegel

Best-selling author

Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die

ABOUT THE AUTHORS



JEFF DEAL is the Vice President of Operations for Elder Research. Working out of the firm’s corporate headquarters in Charlottesville, Virginia, he oversees operational, contractual, and financial matters. Drawing on his more than twenty-five years of management experience in business and government, he regularly helps organizational clients clarify and attain their data analytics goals.

A frequent speaker on the subject of organizational challenges to meeting data analytics goals, Mr. Deal is the program chair for the annual *Predictive Analytics World – Healthcare* conference, which annually attracts leading analytics professionals in the healthcare industry from around the country. He holds a Master of Health Administration degree from Virginia Commonwealth University in Richmond, Virginia, and a Bachelor of Arts degree from the College of William and Mary in Williamsburg, Virginia, where he was a member of the wrestling team. Jeff and his wife, Jennifer, have four children. In his spare time he enjoys hiking, reading, and an increasing amount of recreational travel with his wife, now that their kids have all moved out of the house.



GERHARD PILCHER, Chief Executive Officer for Elder Research, is responsible for the firm’s northern Virginia office. He has more than thirty years of industry and consulting experience with commercial businesses and government institutions in the United States and abroad. His specialties include fraud detection, financial risk management, and healthcare outcomes.

Mr. Pilcher earned a Master of Science degree in analytics from the Institute for Advanced Analytics at North Carolina State University in Raleigh, North Carolina. He is an adjunct faculty member in the math and statistics masters degree program at Georgetown University, and a regular instructor at the SAS Business Knowledge Series course, “Data Mining: Principles and Best Practices.” Gerhard currently serves on the advisory boards of the Institute for Advanced Analytics and the Masters in Science in Business Analytics program at George Washington University. Gerhard and his wife, Denise, have two children. In his spare time, he especially enjoys outdoor activities, including mountaineering and trail running.

ABOUT ELDER RESEARCH, INC.

ELDER RESEARCH, Inc. is a recognized industry leader in the science, practice, and technology of advanced analytics, with vast experience in data transformation and model construction. Founded in 1995 by Dr. John Elder, the company has helped many government agencies and Fortune Global 500 companies solve real-world problems in diverse industry segments.

Elder Research's areas of expertise include data science, text mining, data visualization, scientific software engineering, and technical teaching. By combining the business domain expertise of its clients with its own deep understanding of advanced analytics, the firm creates teams that can extract actionable value from data and transform data, domain knowledge, and algorithmic innovations into world-class analytic solutions.

Corporate Headquarters:

300 West Main Street, Suite 301
Charlottesville, VA 22903
434.973.7673

Additional Offices:

Washington, DC
2101 Wilson Blvd., Suite 900
Arlington, VA 22201

Raleigh, NC

14E Peace Street
Raleigh, NC 27604

Baltimore, MD

839 Elkridge Landing, Suite 215
Linthicum, MD 21090

FOREWARD

For predictive analytics to work, two different species must cooperate in harmony: the business leader and the quant. In order to function together, they each have to adapt. On the one hand, the quant needs to attain a business-oriented vantage. And on the other, the business leader must navigate a very alien world indeed. This book helps with that second bit.

Bridging this gargantuan divide is worth the effort. Take, for example, a tax fraud detection story worth ten digits (covered in the Introduction). Elder Research, Inc., the consultancy that spawned this book, delivered predictive models to the IRS that increased the agency's identification of a certain type of tax fraud by a factor of twenty-five. This saved the Feds billions (with a *b*).

This success exemplifies a widely applicable paradigm. Across commercial and government sectors, predictive targeting achieves a multiplicative improvement to broad scale operations (albeit often a single-digit multiplier rather than that whopping twenty-five-fold improvement). In addition to deciding which tax returns to audit, predictive models determine which customers to contact for marketing, which debtors to approve for increased credit limits, which patients to clinically screen, which employees to woo away from quitting, which persons of interest to investigate, and which equipment to inspect for impending failure.

THUS, DATA SCIENCE EARNS ITS STATUS AS HOT, LUCRATIVE, AND SEXY.

Thus, data science earns its status as hot, lucrative, and sexy. This is the Information Age's latest evolutionary step, technology that taps data to drive decisions more effectively. It's the very act of scientifically optimizing resource allocation for...just about all processes. Various outlets have dubbed *data scientist* as the best, most in-demand,

and even "sexiest" job. And if you haven't heard, data is the new oil. Industry research forecasts that demand will continue to grow and estimates the global predictive analytics market could reach as high as \$9 billion by 2020.

To capture this value, you must construct a durable bridge across the quant/business culture gap. The core technology—which learns from data to predict—is only half of the trick. Deploying it is more than just a technical

process—it's an organizational process. Existing business operations must change by way of implementing analytics. It's no longer business as usual; science now drives the enterprise's primary decisions and actions en masse. In this sense, data science is intrinsically revolutionary.

As a result, the greatest pitfall that hinders analytics is to not properly plan for its deployment. For each analytics initiative, it's critical to build a pathway from the get-go that will lead to integration. This requires bridging the cultural gap. It takes the socialization of buy-in: Line of business staff must agree to make big changes. To that end, they must learn what a predictive model does for them and they must be willing to put their faith in it.

That doesn't always work out. With refreshing frankness, this book reveals an Elder Research study of their own early client projects that showed a full third of projects fail to attain business results, despite 90 percent attaining technical (analytical) success. The difference is often whether the organization actually implements the fruits of analysis.

No guts, no glory. With inertia, resistance to change, and a lack of confidence as primary impediments, there's no more eye-catching antidote than Dr. John Elder's legendary willingness to put his money where his mouth is. Before founding Elder Research, John once invested all his own personal assets into his own predictive stock market trading system, in response to hesitancy on the part of his client to move forward (I recount this story in detail in the book *Predictive Analytics*). And in a story from Chapter 6 of this book, *Mining Your Own Business*, John doubled down against a major credit card company, betting Elder Research could beat their established analytical methods to model credit risk. If Elder Research failed, they'd cut their service fee in half, but if they won, the cost would double (yet, in the latter case, the company would gain enormously from the improved predictive model). This tactic served handily to move the project forward.

Don't worry. When inertia hinders progress, you don't necessarily need to take the dramatic approach of wagering your own money. There are other options among the established best practices for managing predictive analytics initiatives. It's largely about educating the organization and opening discussions to understand the concerns of skeptics.

Unfortunately, convolution and the appearance of arcane complexity threaten to extinguish a newcomer's excitement about the potential value. This might leave the person feeling compelled only by the pressure that

comes from hype: “Everyone’s doing it!” Let’s nip that in the bud right now. Predictive analytics’ value is simple and concrete: it helps run operations more effectively by way of predicting behavior, i.e., the outcome for each individual consumer, employee, healthcare patient, or suspect. These predictions are each just numbers, aka scores or probabilities. Since they directly drive decisions, by definition they are the most actionable deliverable you can get from analytics. One need only learn a limited bit about the (fascinating) “rocket science” that generates these predictive scores to integrate them and realize their value.

Regrettably, today’s tremendous data hype does not always relay this value proposition or any specific value proposition at all. The pervasive buzzwords *big data* and *data science* enthusiastically remind us there is value to be had, but do not refer to any particular technology or approach. These terms are general catch-alls for “doing smart things with data.” They really have no agreed upon definition beyond that, although they do allude to a vital cultural movement led by thoughtful data wonks. *Big data* is nothing more than a grammatically incorrect way to say “a lot of data” (like saying “big water” instead of “a lot of water”). *Data science* is a redundant term, since all science involves data; it’s like saying, “book librarian.” In Chapter 2, the authors of this book delve deeper with solid coverage of the extensive taxonomy of terms and technology.

**THIS BOOK IS
DIFFERENT.
JEFF DEAL
AND GERHARD
PILCHER WROTE
IT TO SERVE
THE MUCH
NEGLECTED
OTHER SIDE OF
THE COIN: YOU,
THE BUSINESS
LEADER.**

In a field propelled largely by data nerds, it may come as no surprise that most books serve the hands-on quant. Those books dive into the technical practice. After all, for a quant, the technology and software tools are much more tangible and easy to define than the more elusive, “human” arena of organizational processes and project management. As a natural-born geek, I know from personal experience.

This book is different. Jeff Deal and Gerhard Pilcher wrote it to serve the much neglected other side of the coin: you, the business leader. It delivers the two ingredients you need for success: 1) an understand-

ing of the technology so you can speak the quant's language and 2) a guide to analytics management best practices, including how to build your analytics team and avert the most costly pitfalls.

In this book, the authors hand over an unmatched treasure trove of anecdotes accumulated at the firm from which they hail as executives—Elder Research, the most widely experienced data consultancy in North America.

In contrast to vendors of analytics software, which traditionally are less likely to disclose the rough patches and challenges often experienced with analytics projects, Elder Research is an industry leader in deployment, across analytics software solutions and across sectors. The firm has accumulated a highly diverse portfolio of experience working with all major industry verticals as well as many government organizations, including defense, intelligence, and civil government.

With this book, Deal and Pilcher extend Elder Research's track record of thought leadership and industry education. This firm—and the two authors in particular—regularly contribute to the conference series I founded, Predictive Analytics World. They head up two of its annual events, PAW for Government and PAW for Healthcare, and also provide various acclaimed presentations and training workshops at several other PAW events each year. If you have the opportunity to meet them in person, take it—but before you do, assimilate the wisdom they've labored to set forth in this book.

ERIC SIEGEL, PH.D., founder of Predictive Analytics World and author of *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (revised and updated edition)

INTRODUCTION

Data mining and predictive analytics have been much in the news of late. In spite of all the hype—perhaps partly *because* of it—there’s a lot of confusion about what these powerful decision-making tools are, how they function, and how they can best be utilized.

That’s why we wrote this book. It’s an easy-to-read, practical primer for C-level to mid-level executives about how to harness the power of these state-of-the-art technologies to increase organizational effectiveness. If you’re a leader in an enterprise-level organization or a consultant to such organizations, you will find this book to be a useful introduction to the field and a valuable resource you’ll refer to again and again.

When John Elder started our firm back in 1995, data analytics was in its infancy. We like to say that we were data scientists before it became cool. As early visionaries, we not only witnessed this revolution in management practice, we helped to create it. And we’re still doing that today! Over the past two decades, Elder Research, Inc. has served hundreds of clients in industry, government, and academia. We’ve acquired a wealth of information about what works and what doesn’t, and in the pages of this book we share this inside knowledge with you. Below are some of the topics included:

- How to foster an analytic culture
- When and where data analytics can be most useful
- How the analytic process works from beginning to end (Hint: It involves more than analytics.)
- How to organize, staff, and manage data analytics initiatives
- How to build, implement, and evaluate data analytics models (an overview)
- When and how to hire analytics consultants
- How to approach the decision of what analytics software to purchase, if any
- How to avoid the most common data analytics mistakes
- How to integrate analytics into everyday workflows

Data analytics has been called the most powerful decision-making tool of the 21st century. Even though it has come of age only within the past twenty years, thousands of businesses, governmental agencies, and nonprofit organizations have already used it to dramatically increase productivity, reduce waste and fraud, enhance quality, improve customer service, boost revenues, evaluate applicants, optimize strategies, combat crime and terrorism, and solve a host of other tough challenges. Following are three actual cases from our company's files that help illustrate its value.

CASE 1

A worldwide provider of computer technology and consumer electronics knew that some service providers were submitting fraudulent claims and pocketing the reimbursements. This showed up in the simplest way, for example, when some of the same product serial numbers popped up again and again, but management did not know the extent of the problem, and they were oblivious to other more sophisticated scams. Our data scientists were able to discover several types of fraud schemes, determine the scope of the problems, identify the culprits, and develop corrective procedures. As a result, the company saved an estimated \$20 million in the first six months of our consulting engagement and more than \$75 million in the first five years.

OUR DATA SCIENTISTS WERE ABLE TO DISCOVER SEVERAL TYPES OF FRAUD SCHEMES, DETERMINE THE SCOPE OF THE PROBLEMS, IDENTIFY THE CULPRITS, AND DEVELOP CORRECTIVE PROCEDURES.

CASE 2

The IRS needed to improve its models to identify fraud related to a type of tax refund. A serious data mining project by a team of companies overhauled the data features and radically improved the model, which resulted in a 25-fold improvement in positive identification of fraud. The model and project are credited with saving over \$7 billion in its first several years of use.

CASE 3

A regional provider of telecommunications services wanted to reduce “churn,” which is the industry’s term for customer account closures. Data mining work assisted management in determining the major causes of churn, and predictive analytics helped identify which customers were most likely to churn in the future. Developing solutions based on this data allowed the company to add approximately \$1 million per year to its bottom line. In fact, these solutions currently generate more than enough savings each *month* to pay for the *entire* multi-month consulting engagement.

A management revolution is underway in the world of business and government. In the years ahead, the most successful organizations will be analytically competent. This book will help you gain the knowledge, vision, and passion you need to be on the cutting edge of this revolution.

HOW TO USE THIS BOOK

Mining Your Own Business provides an easy-to-read overview of data mining and predictive analytics for organizational leaders who want to know more about these powerful tools. It will also serve as a reference for those who want to develop an analytic capability in their organization.

Below is a summary outline of the book, with some suggestions about the level and type of leadership involvement required at each stage of the analytics process.

ESTABLISHING A FOUNDATION AND VISION FOR ANALYTICS

This stage is critical to everything that follows. From the beginning, leaders need to invest enough time understanding the business and its mission, so they are able to identify and articulate to the organization the opportunities that analytics can address. They must prioritize goals and develop plans for the analytics initiatives, while taking into consideration the organization's constraints on funding, staffing, data availability, consultants, tools, and deployment.

▶ See **Chapters 4, 5, 6, and 12** for a discussion of these topics.

**LEADERS MUST
LEARN HOW TO
DEAL WITH THE
CHALLENGES
OF CHANGE,
SO THEY CAN
SUCCESSFULLY
LEAD ANALYTICS
INITIATIVES WITH
CONFIDENCE,
PATIENCE, AND
PERSEVERANCE.**

FOSTERING A CULTURE OF ANALYTICS

The extraordinary results produced by data analytics usually require the organization to undergo disruptive changes. Leaders must learn how to deal with the challenges of change, so they can successfully lead analytics initiatives with confidence, patience, and perseverance.

▶ *Managing change is discussed in some detail in **Chapters 3 and 13.***

MONITORING ANALYTIC PROJECTS

Monitoring Analytic Projects Early on, analytics leaders should devote considerable energy to understanding the business, defining the objectives of the

initiative, and establishing measures of success. Once the data preparation and modeling begins, they need to monitor the progress of the project to ensure it meets the organization's objectives.

Leaders also must consider implementation of the model, because implementation constraints can impact the choice of modeling techniques and even precipitate disruptive changes to the organization's established business processes. Failure to initially budget sufficient finances and time for implementation can turn a valuable piece of software (the analytic model) into "shelfware" (a model unused by the business).

▶ *See **Chapters 7, 8, and 12** for information on leading and monitoring analytics projects. **Chapters 9 and 10** provide some technical details about modeling that will help leaders ask the right questions when hiring consultants and when assessing the quality and progress of projects.*

BUILDING CONFIDENCE IN ANALYTIC RESULTS

It is the leader's responsibility to assess the process used to validate the model. An analytic model may be validated by asking the essential question, "Are the results it produces better than the results currently produced by other methods?" Sometimes the "other method" may be nothing more than "gut feel" or "educated guesstimates."

▶ *This topic is addressed in **Chapter 11**.*

PUTTING ANALYTIC MODELS INTO OPERATION

The leader plays a critical role in defining the business objective, which will drive the ultimate operationalization of the model. The many different implementation possibilities may range from simply running the model on an annual basis to creating a set of enterprise-level applications that incorporate analytic results into a business process. Outside consultants (if involved) typically pass the leadership role to the in-house team at the beginning of the implementation phase.

▶ *We address this topic in **Chapter 12**.*

ADDITIONAL RESOURCES

We've listed some additional resources for further study on the book's website: www.MiningYourOwnBusiness.com.

Chapter 3

LEADING A DATA ANALYTICS INITIATIVE

Leading a data analytics initiative is in some ways similar to gardening. Just as the first step in gardening is deciding what plants to grow, the first step with analytics is deciding what problems to solve. Begin by identifying a narrowly defined problem that is widely acknowledged within the organization as a pain point. One example might be customer or employee attrition.

As a data analytics goal, however, “reduce attrition” is overly broad. A better starting point might be “to identify characteristics associated with attrition.” This more narrowly defined problem will be easier to solve, and it will still produce useful insights about attrition. A common misstep when first starting an analytics initiative is to take on more than the organization is capable of handling.

Organizations of any size have many problems data analytics can help solve. In picking your initial project, consider what problems keep you awake at night. What forces are impacting your organization that you need to understand better? What issues would you very much like to resolve? To ensure success, leaders should view the initiative as a vital step toward attaining the organization’s overall goals. All involved in the project should buy into its success, and they should be willing to act on the insights the analytics will uncover. The project should be practical, with adequate funding and access to data. Especially at the beginning of the analytics initiative, it’s essential to ask the right questions.

In the first chapter, we listed a range of problems our clients have addressed. Most of these types of projects could serve as a good starting place for your organization, as long as the scope is clearly defined. Some beginning projects that have worked well for our clients are reducing fraud, reducing customer turnover (churn), predicting costs, and improving production quality. Limit the focus of your effort. If your issue is fraud, for example, begin by identifying and reducing one particular type of fraud, rather than going after every form at once.

A primary purpose of your initial project should be to prove the value of data analytics and get people on board about its potential benefits. Obtaining buy-in is key, because analytics ultimately will change the way people in the organization make decisions. You must be prepared to lead change if you are going to thrive as an analytically-driven business. And to successfully lead change, you must gain commitment from the people who will be affected by producing useful results at an early stage.

View this initial data analytics project largely as a sales tool for future full-scale initiatives. Use it to demonstrate to all levels of management that the benefits of data analytics merit the required investment of time, money, and emotional energy.

Data analytics initiatives usually lead to significant changes in the organization's procedures and culture. Fundamentally, an analytic result should impact an organization's decision-making process. Most of us, at some point in our careers, have experienced difficulty implementing or adjusting to organizational change. There are many excellent books about leading change and overcoming human resistance to it, so we'll simply mention a few aspects specifically related to analytics.

For example, when implementing a model to reduce credit card fraud, consider the fundamental changes that have to be put into operation in order to process credit card transactions differently. The system has to score hundreds of millions of transactions per day and keep a record of recent, related transactions. When a transaction exceeds established risk thresholds, the system must automatically suspend approvals. And even before the system can become operational, the credit card company has to retrain all call center attendants and provide them with a set of tools to assist owners of suspended credit cards.

**YOU MUST BE
PREPARED TO
LEAD CHANGE IF
YOU ARE GOING
TO THRIVE AS AN
ANALYTICALLY-
DRIVEN
BUSINESS.**

Changes such as these can be uncomfortable, and they are often controversial. That's why it's crucial to sell the benefits of data analytics and gain buy-in during the early stages of the project.

Occasionally an organization with no specific purpose in mind asks us

to come in and help it build a predictive analytics model. The executives assume we can provide a magical “black box.” They expect us to dump their data into it, turn the crank, and generate insights that will enable them to transform the way they run their business. Occasionally we can produce some extraordinary results from that sort of non-directed approach, but we strongly discourage it.

If there are no burning issues keeping people in your organization awake at night, and if no one is asking for the results you’ll create, the project will probably turn out to be a waste of time and money. That’s exactly what happened with one large government contractor who hired us to build the organization’s analytics capability.

When we asked the senior executive of this organization how he intended to use the insights the model would produce, he said he didn’t know, but that he was confident several units in the organization would find the information useful. We cautioned him against imposing a project from the top down, and we urged him to find a sponsor or client within the organization who wanted a particular problem solved. We went on to explain that committed stakeholders are much more willing to contribute to the project by providing subject-matter experts and cooperating in other ways.

Unfortunately, this executive insisted on proceeding with his “if we build it, they will come” approach, and our fears materialized. Even though the model we built produced good results, no one within the organization was interested in them. For its considerable investment of time and money, this client merely ended up with a solution in search of a problem.

STARTING SMALL

Some companies new to data analytics try to push ahead too fast. Instead of investing \$75,000 or so in a modest initial project, they rush out and hire two or three people who have some experience with analytics, spend \$500,000 on software, and announce that the company is now “datadriven.” However, pursuing a data analytics initiative without proper planning and organizational buy-in is like purchasing an expensive piece of home exercise equipment without sufficient commitment. The equipment may seem exciting at first, but without a dedicated regimen, it will soon end up sitting idle in the basement or serving as a clothes rack in the corner of the bedroom.

Several years ago the managers of a very large company asked us to help them with their vision for using data analytics as a vehicle for transforming the entire healthcare industry. After our team of four people met with their team of twelve people for two days, it was clear that the organization's vision was too grandiose to get off the ground. It was as if they were trying to reach the moon with a hobbyist's model rocket. After two days of meetings, the potential project collapsed under the weight of its unrealistic goals.

In contrast, another of our clients, one of the largest insurance companies in the United States, did everything correctly. When we came in for the initial kickoff meeting, the leader of the project had already assembled the key SMEs, executives, IT people, and other stakeholders.

The project this company presented to us involved a widely acknowledged, well-defined pain point associated with a particular line of insurance. Its clear focus and narrow scope led to highly productive meetings and substantial buy-in.

Although committing all of these people to a day and a half of meetings was expensive, the investment paid off. The initial project was very successful, and since then the company has applied data analytics to other problems associated with this same type of insurance. In the future, they plan to expand data analytics into other lines of their business.

Something else impressed us about this client. We were delighted to find that the person responsible for data security was very forward-thinking. She was determined to do everything possible within the law to make the information our initiative required available. In our experience, far too many data security people are afraid to share any information that contains personal data about customers or clients. Fortunately, this person put forth the extra effort to get us the data we needed, without violating customer

**PURSUING A
DATA ANALYTICS
INITIATIVE
WITHOUT PROPER
PLANNING AND
ORGANIZATIONAL
BUY-IN IS LIKE
PURCHASING
AN EXPENSIVE
PIECE OF HOME
EXERCISE
EQUIPMENT
WITHOUT
SUFFICIENT
COMMITMENT.**

confidentiality or the law.

EXAMPLES OF POOR VS. GOOD FOCUS

Let's discuss two cases that demonstrate the importance of focus. The first involves a national lending organization that asked us to create an analytic model for identifying high-risk loans. The firm operated in most of the fifty states, and each state had different types of data, different methods of storing data, and different laws pertaining to the lending business. The company wanted one model for all states that would score potential borrowers on their likelihood of defaulting on their loans.

Although the overall goal was reasonable, we immediately saw that the company would need almost fifty different models. Unfortunately, when we told management that individual models would be needed for each state, they did not take our advice. Due to this and other organizational problems, the project never got off of the ground.

A more positive outcome resulted when a prestigious post-graduate school at a large university asked us to help them identify which applicants were likely to accept an offer of admission, if extended. The school provided us with several years of applicant data that had been scrubbed of personal identifying information. It included all of the information pertinent to the admission decision, such as the name of the undergraduate college the applicant had attended, the applicant's major as an undergraduate, and the amount of financial assistance the institution was offering the applicant. For the training data, we also knew the outcomes (i.e., which applicants had received offers of admission, and which offers had been accepted).

The university decided to focus the model on a very simple question: "If we make an offer of admission to a student, what is the probability that this student will accept the offer?" We could have built a model that would answer other questions as well, such as, "If we admit a certain applicant, what is the probability that this applicant will eventually graduate?" Or, "If we admit a certain student, into what quadrant will this student likely fall in the class rankings?" To their credit, the school resisted the temptation to broaden the initial focus. Their narrowly defined target made each aspect of the work easier and more efficient. This project was a big success and a pleasure to conduct.

CULTIVATING THE CULTURE

Returning to our gardening analogy, after deciding what to plant, the next step is to cultivate the soil. Similarly, the second step in data analytics is to cultivate the organizational culture. This entails developing the analytic plan, setting goals, and gaining buy-in from the people who will be involved.

Because a data analytics initiative will draw on knowledge, services, and resources from multiple areas, commitment to it must be broadly shared within the organization. All participants need to view the project as an important component of the company's vision. A project with a weak commitment will likely wither and die.

A predictive analytics project we conducted for a major government agency revealed that the satisfaction of customers increased when response times to their complaints decreased. This insight surprised no one, but the project did arouse attention when it identified a point beyond which further reductions in response times were not worth the cost. In other words, the agency's customers weren't noticeably happier when the customer service response times were cut from forty-eight hours to twenty-four hours. Based on this analysis, we demonstrated that responding less quickly could save the agency about \$2 million dollars per year, without any significant loss in customer satisfaction.

Unfortunately, more than two years have passed, and this agency has yet to optimize its response times in accordance with our recommendation. We're always disappointed when we see a predictive analytics program with a large potential benefit discontinued in midstream, but it happens quite a lot. As with most other significant organizational changes, implementation of data analytics recommendations will usually require adjustments to the company's culture and processes. These changes are never easy, and sometimes implementation is further hindered by corporate politics or bureaucratic inflexibility.

In contrast, the leaders of another organization we worked with did an especially fine job of laying the groundwork for cultural transformation. One of their key moves was to ask a few potential users of the results to serve on the initial installation team as subject-matter experts for their particular areas of responsibility. Although these individuals were not data-driven people by nature, their exposure to the data analytics process turned them into believers. When they saw the potential benefits that the

data-driven models would produce in their own areas, they became very effective cheerleaders. Other potential users of data in the organization jumped on board. Soon, virtually everyone had bought into the initiative.

MANAGING A DATA ANALYTICS INITIATIVE

With a data analytics program, a small success serves as the foundation for greater success. As with a garden, long-term success takes commitment, patience, and on-going tending. Just as a gardener must water and fertilize the garden until the plants appear, the leaders of a data analytics initiative must provide encouragement and guidance until the insights surface and appropriate actions are undertaken. And just as a gardener must pull weeds that would crowd out the plants, the leaders of a data analytics initiative must guard against distractions that would undermine the project's success.

Because the results of a data analytics project can take quite some time to manifest, staying the course can require considerable patience and perseverance. As an organization continues to invest time and money into the initiative, some leaders may get anxious or even fearful. "What if this project doesn't pan out?" they may start thinking. "Are we simply pouring money down the drain?"

BECAUSE THE RESULTS OF A DATA ANALYTICS PROJECT CAN TAKE QUITE SOME TIME TO MANIFEST, STAYING THE COURSE CAN REQUIRE CONSIDERABLE PATIENCE AND PERSEVERANCE.

Sometimes other people will suggest other problem-solving techniques, and leaders may be tempted to divert funds from the data analytics initiative to alternative approaches that may appear to be faster and cheaper. Leaders should continually remind themselves and all involved that quick fixes rarely provide effective long-term solutions to complex problems. Courageous, positive leadership is necessary to ensure that the data analytics initiative stays on course until successes begin to manifest. At the same time, the analysts need to keep those working on the initiative encouraged with reports of early findings.

THE EXPERIENCES OF A MOBILE PHONE SERVICE PROVIDER

The management of a regional provider of mobile phone services asked us to build an analytics model to predict which of their customers were most likely to “churn” (i.e., switch to a competing carrier). Our client planned to give these predictions to call-center operators, who would then telephone these high-risk customers prior to contract expiration and explain the advantages of renewing their current contract.

We had confidence in the model we built, but its results, during the first several weeks of operation, were disappointing. Even though call-center operators were telephoning the customers we had identified as high risk to try to persuade them to extend their contracts, churn remained high. Management wondered if their rather significant investment in predictive analytics had been a mistake. Even we were starting to feel perplexed!

Upon further investigation, however, we discovered that the model was working fine. The problem was with the scripts that call-center operators used when making calls. It turned out that when operators were unable to reach the customers identified as high risk, they were leaving voicemail messages that their service contracts were about to expire. These messages had the unintended negative effect of alerting these customers that they would soon be free to change carriers without penalty. This resulted in more churn, not less, because these calls essentially prompted customers to change carriers when they otherwise might not have thought about it. As the saying goes, it would have been better if the call-center operators had “let sleeping dogs lie.” Our discovery about how call-center procedures were actually promoting churn created the first of a series of “aha moments” for the company.

When management told operators to forego voicemail messages, churn decreased dramatically. A comparison of the actions of potential churners who received a phone sales pitch to the actions of a control group that did not have the benefit of the sales pitch showed that our initiative was saving the company more than 200 customers per month. This quantitative evidence of the value of predictive analytics created the second “aha moment” for the client. These results were especially gratifying to our client, because at this time, the company was operating at a competitive disadvantage in the marketplace. The newly released iPhone was dominating the market, and at that time our client was not authorized to sell it.

A few months later, when the company became an authorized distribu-

tor of Apple products, management decided to deemphasize this initiative. They figured that churn wouldn't be an issue, because customers no longer had to go to a competitor to buy the iPhone. Under pressure to divert some of the data analytics resources to other initiatives, management cut the call-center support for the analytics campaign by two-thirds.

The results were disastrous. The company's overall churn increased dramatically, so that our campaign was now preventing zero churns per month. When management reestablished the analytics initiative to full strength, churn savings shot up past 200 to more than 500 customers per month. This third "aha moment" convinced the client of the value of predictive analytics.

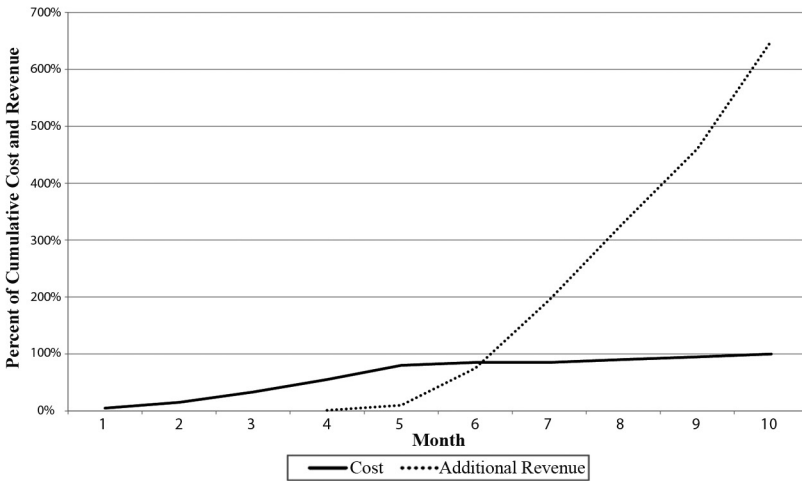


Figure 3-1: Cumulative Percentage of Cost and Revenue

Figure 3-1 illustrates the costs and benefits of the first ten months of this program. The solid line depicts the company's cumulative investment in the project, and the dotted line represents the cumulative additional revenues produced by the initiative. Note that management had to invest funds into the project for four months before any new revenues began to materialize. As we have said, leaders must exercise patience, perseverance, and sometimes even courage while waiting for results of data analytics initiatives to manifest.

This client did not give up, and in the fifth month the payoff started to

appear. Initially, the return on investment was small, for reasons explained above. After the company modified its call-center procedures and added the iPhone to its product line, the additional savings produced by the model accelerated rapidly. At month six, the company reached a breakeven point, where the cumulative additional revenue generated by the initiative had fully offset the cumulative amount invested in the project.

By this time, the company had invested about 95 percent of the total amount of funds that would be required, and the additional cumulative revenues were rising rapidly. By month ten, the additional revenues produced by the model were 6½ times greater than the cumulative cost of the project. On an annual basis, the return on investment was above 750 percent and still rising! In fact, with each passing month, the business generated enough additional savings to pay for the entire cost of the investment in analytics. Today the company continues to reap substantial benefits from this model, and it is expanding the application of analytics to other areas of its business.

BRYAN'S STORY

Part 1: Launching a Data Analytics Initiative

Starting in this chapter, Bryan Jones will share his experiences as the leader of a successful data analytics project. We will include additional installments of his story in subsequent chapters. At the time of this project, Bryan was deputy assistant inspector general for analytics for the U.S. Postal Service Office of Inspector General. Here is the first installment of his experience in his words.

I'm somewhat surprised that I got involved with analytics in the first place. I was never good at math, and it took me two years to get through Algebra II. But I could translate data into terms that business people understood and found useful.

I believe my success as a leader of data analytics projects was due largely to the business and personal relationships I had built over eight or nine years with the inspector general and with a lot of the directors, managers, and staff. Even if people didn't

understand analytics or care a lot about it, they would listen to me and give me the benefit of the doubt, because of the strength of our relationship and the credibility I had earned with them. I would say something like, “You guys are the experts. If you give us a good problem to solve and send us somebody to work on it, we’ll try to put something together. When we’re done, you can tell us if it’s valuable.” I knew from my background as an auditor that we had a good product. My sincere desire was to help others see its value and benefit from it.

LEADERSHIP IS KEY

A study our firm conducted of the projects we had completed in our first decade since our founding in 1995 showed that 90+ percent had been technical successes, but only 65 percent of those had been implemented; that is, only two-thirds had been business successes.⁹ Many causes can contribute to a lower business success rate, but the biggest one is a lack of organizational commitment to implementation. Unfortunately, many organizations simply are not willing to operationalize the recommendations that the analytical work provides, even when it’s obvious that these recommendations will lead to significant improvements.

Why do so many organizations invest significant amounts of time and money in a data analytics project and then fail to implement the resulting recommendations? By this stage, they’ve paid all the costs, and they have proven returns on out-of-sample data. Without implementation, however, they realize no gain.

⁹ Studying our past – data mining our data mining – proved very fruitful. By identifying the major non-technical obstacles to putting a new model and process into production, we were able to avoid most roadblocks in the future and get our production rates up substantially in the following decade.

We believe the two major reasons are the absence of strong leadership and a lack of buy-in by key decision makers. (A minor reason is failure to understand the results, which is a failure on the part of both analyst and client.) To implement data analytics recommendations, organizations often must develop new policies and procedures, change long-standing processes, retrain personnel, and even transform corporate cultures.

INCREASE THE PROBABILITY OF SUCCESS BY INVOLVING KEY STAKEHOLDERS FROM THE BEGINNING.

es, retrain personnel, and even transform corporate cultures.

Increase the probability of success by involving key stakeholders from the beginning. A data analytics project is likely to fail if you conduct it without involving the stakeholders, and then upon completion tell them, “Here’s what the data shows. Now make use of it.”

Changing a corporate culture is never easy. That’s why successful predictive analytics initiatives demand strong leadership from one or more “champions” who are enthusiastically committed to analytics and who command sufficient respect within the organization to enlist the commitment of others. We’ll close this chapter by relating a few stories about what has worked well and not so well with some of our clients.

A PARADE OF CHAMPIONS AT A FEDERAL AGENCY

In a successful project we conducted for one of the larger federal oversight agencies, the head of the agency was the visionary champion. He established the cultural expectation that the organization would become data-driven, and he used his budget authority to make sure the funds would be available to pay for the necessary consultants, software, and internal services.

A short time later, another champion who was quite analytics-minded arose from a lower level of the organizational ladder. He identified a specific problem that needed to be solved and helped to convince people that data analytics was the best approach. About two years later, a third champion emerged. This key person, a very practical leader, became the quarterback and cheerleader who led the first project to successful completion. Each champion was able to build on the momentum and foresight of the previous one. Because all three operated at different levels in the organization, they

could communicate strategically or tactically as that level demanded.

A LACK OF LEADERSHIP AT A FINANCIAL FIRM

Another case from our files didn't turn out so well. After a preliminary analysis of data, we submitted a proposal to a Fortune 500 financial firm that showed that for an investment in data analytics of about \$500,000, this firm could reduce its overseas workforce by about 75 percent. This promised to save tens of millions of dollars annually, with no adverse effects on operational effectiveness. Management could allow these savings to flow through to the bottom line, or perhaps even better, they could use them to train the company's foreign workforce to collect and verify data needed for new products it wanted to offer.

Three years have now passed, and the company has failed to act on this recommendation. Why? It's not an issue of money. Yes, \$500,000 is a significant investment, but the potential return on this investment is huge. Technical issues aren't the problem either. Management knows that predictive analytics is a proven science, and they agree that the preliminary study was a big success.

The fundamental problem is lack of leadership. Management started off on the wrong foot by failing to involve key stakeholders. This led to less than optimal buy-in. When the project got underway, various departments dug in their heels. The company's IT department, for example, viewed the initiative as a threat. They were afraid it would take work away from them and possibly make them look bad. Along with other units in the organization who had similar feelings, they sold management on the idea of doing the project internally. Since the company lacks the analytics capability to do the job right, to this day the work remains unfinished and ineffective, and the savings remain unrealized. This type of corporate paralysis is not unusual, especially when the decision involves many different stakeholders within an organization.

THE EFFECT OF DIFFERENT LEADERSHIP STYLES AT A GOVERNMENT AGENCY

Another government agency we have worked with has had very good success with predictive analytics in one area of the organization and much less success in another. The difference is due almost entirely to the leader-

ship styles of the key people involved. In the first case, the key person is a true champion of data analytics. He believes in its potential, and he's very willing to implement the recommendations it produces. In the other case, the key person is extremely cautious about data analytics and hesitant to act on the insights it generates. He sees it more as a drain on resources than as a generator of value. Unfortunately, this mindset has turned out to be a self-fulfilling prophecy.

BOLD LEADERSHIP REQUIRED

The insights produced by a predictive analytics process are worthless unless acted upon. Unfortunately, many organizations lack the courage and determination to implement the resulting recommendations, especially when they run counter to conventional wisdom, or when they demand substantial changes to long-standing procedures. We humans are more strongly influenced by our biases than we generally realize. Learning to trust and utilize data instead of relying on past experience and “gut feel” can be as challenging to business leaders as learning to fly an airplane by instruments rather than by sight can be to beginning pilots.

The difficulty of implementing analytics recommendations is captured in the fascinating book (and movie) *Moneyball*, which describes how Billy Beane, the general manager of the Oakland Athletics, had to battle seasoned baseball veterans and long-standing baseball traditions in order to install a data capability for his team. Even his own team manager resisted his efforts. To champion this transformation, Beane had to be the visionary, cheerleader, encourager, teacher, cajoler, and drill sergeant—essentially, he had to play every role except that of the analytic model builder. Such leadership is demanding and often lonely, but in this case Beane successfully transformed the culture and made a good baseball team a whole lot better, ultimately generating a huge reward. He serves as an excellent role model for aspiring leaders of analytics initiatives.

CHAPTER NUGGETS

- Most organizations, of any size, have problems data analytics can help solve.
- In picking your initial project, consider what problems keep you awake at night.
- Obtaining buy-in from fellow stakeholders is key because analytics ultimately will change the way people in the organization make decisions.
- Data analytics initiatives usually lead to significant changes in the organization's procedures and culture. These changes can be uncomfortable or even controversial, so it's crucial to sell the benefits of data analytics and gain buy-in early on.
- Because a data analytics initiative will typically draw on knowledge, services, and resources from multiple areas, commitment to it must be broadly shared within the organization.
- With a data analytics program, a small initial success serves as the foundation for greater success.
- Because changing a corporate culture is never easy, successful predictive analytics initiatives demand strong leadership from champions who are enthusiastically committed to analytics and who command sufficient respect within the organization to enlist the commitment of others.
- The insights produced by a predictive analytics process are worthless unless acted upon. Unfortunately, many organizations lack the courage and determination to implement the resulting recommendations, especially when they run counter to conventional wisdom, or when they demand substantial changes to long-standing procedures.

MORE ABOUT MINING YOUR OWN BUSINESS

If you enjoyed Chapter 3 and would like to learn more about the remaining chapters of *Mining Your Own Business* you can view the complete table of contents on the following pages or visit **www.miningyourownbusiness.com**.



CONTENTS

Foreword.....xiii

Introduction..... xvii

How to Use This Book..... xix

Introduction and Overview

1. Empowering the Decision Makers.....1

- Hunting for Needles in Haystacks.....1
- Breaking the Mind Barrier2
- A World of Applications4

2. Clearing Up the Confusion7

- Ten Levels of Analytics.....8
- Four Categories of Modeling Knowledge.....14
- Supervised vs. Unsupervised Learning.....16
- Levels and Advanced Data Types18

The Analytic Organization

3. Leading a Data Analytics Initiative.....21

- Starting Small.....23
- Examples of Poor vs. Good Focus.....24
- Cultivating the Culture.....25
- Managing a Data Analytics Initiative26
- The Experiences of a Mobile Phone Service Provider.....27
- Leadership is Key30
- A Parade of Champions at a Federal Agency31
- A Lack of Leadership at a Financial Firm.....31
- The Effect of Different Leadership Styles at a Government Agency.....32
- Bold Leadership Required32

4. Staffing a Data Analytics Project.....35

- Individual or Team?36
- Assembling the Team37

- What is a Data Scientist?.....39
- More than Academic Credentials41
- The Most Important Quality42
- Mike Thurber’s Story43
- Building Teams through “Gap Analysis”.....43
- 5. Acquiring the Right Tools.....47
 - A Variety of Techniques and Disciplines48
 - Interface Level of Tools51
 - Sources of Tools53
 - A Word about Open-Source Tools54
 - Tool Trends55
- 6. Hiring Data Analytics Consultants59
 - Discerning Fact from Hype60
 - Evaluating Industry Experience.....61
 - Evaluating Analytics Experience63
 - Finding the Right Consultant63
- The Modeling Process
- 7. Understanding the Data Mining Process67
 - The CRISP-DM Process67
 - Resist the Temptation to Take Shortcuts69
- 8. Understanding the Business73
 - Clarifying Your Objective74
 - Defining the Terminology75
 - Framing the Questions76
 - An Unexpected Finding77
- 9. Understanding and Preparing the Data.....81
 - Understanding the Data81
 - Cleaning the Data.....82
 - “Perfect” Data83
 - Collecting and Preparing the Data84
 - Fostering Cooperation85
 - Governing the Data86

- 10. Building the Model..... 89
 - Inside the “Black Box”..... 89
 - Building an Illustrative Model..... 90
 - Non-Linear Models..... 93
 - Choosing a Model 95
 - Response Surfaces of Predictive Models..... 97
 - The Trade-off between Accuracy and Interpretability..... 98
 - Choosing and Testing Modeling Algorithms 98
 - Dealing with Variance..... 99
 - Model Ensembles 100
- 11. Validating the Model 103
 - Technical Validation..... 103
 - Checking for Mistakes..... 106
 - Checking for Generalization 112
 - Using Experts to Qualify Model Results 112
 - Target Shuffling 114
 - Business Validation 116
- Putting the Model into Practice
- 12. Deploying the Model..... 119
 - Planning and Budgeting for Deployment..... 119
 - Business Processes Are Key 120
 - Example: Finding Taxpayer Fraud..... 121
 - Four Important Questions 122
- 13. Realizing the Transformation..... 125
 - Realizing the Potential..... 125
 - The Tipping Point..... 126
- Appendix
 - Are Orange Cars Really Least Likely To Be Lemons?..... 129
- About the Authors 151
- About Elder Research, Inc..... 152
- Index..... 153

ENDORSEMENTS

“Government and Industry Executives, if you have not been comfortable with buying into a program based on advanced analytics, algorithms, and data scientists, this book will be transformational for you and your program. It captures all the critical elements and decades of experiences into a few clear pages that will light the path for predictive improvements. I will be sharing it with my leadership and program managers. Great job, gentlemen, in making a complex equation simple to follow!”

— **Fred Walker**

Technical Director Counterintelligence, National Security Agency

“Amidst the concerns about the shortage of data scientists, a larger, overlooked obstacle is finding C-, VP-, and director-level leaders who understand enough about advanced analytics to hire, manage, and deploy solutions. Deal and Pilcher have written a practical and insightful ‘primer for executives’ to expertly fill this void.”

— **Dean Abbott**

Co-Founder and Chief Data Scientist at SmarterHQ
Author of *Applied Predictive Analytics*

“Deal and Pilcher have distilled their decades of experience into an easy-to-read book that will benefit any business person dealing with analytics. They keep technical details to a minimum while focusing on the key facts, decisions, and actions that business people need to be successful with analytics. Abundant real-world examples reinforce their practical and valuable advice. Your time reading the book will be well spent!”

— **Bill Franks**

Chief Analytics Officer, Teradata
Author of *Taming The Big Data Tidal Wave* and *The Analytics Revolution*

A Data Mining Resource for Executives

Data mining and predictive analytics are the 21st century's most powerful new management tools. In this practical, easy-to-read guide for organizational leaders, Jeff Deal and Gerhard Pilcher explain

- what data mining and predictive analytics are
- why they are such powerful management tools
- how and when to use them for greatest positive impact across a broad spectrum of industries.

Use this book to gain a quick overview of the subject and as a handy resource to be referred to often. If you're preparing to lead or participate in a data analytics initiative, you must read this book!