

Strategies and Tactics for Successfully Developing In-House Predictive Analytics Expertise

By Daniel Putler & Robert Krider



Predictive analytics, one of the more sophisticated components of business analytics, has achieved wide acceptance among large organizations. Today, as the tools become more widely available and their benefits more obvious, progressively smaller organizations are incorporating predictive analytics into their business processes. Historically, the rapid increase in the amount of data collected internally and easily available externally has been a driving force in analytics adoption. As more organizations adopt the tools and processes, another motivation becomes apparent: organizations that do not use these tools are being left behind by competitors that do. Non-obvious predictions, such as the probability that a specific customer will respond positively to a particular offer, extracted from existing data can inform decisions that provide a large and predictable impact on the bottom line.

Organizations that wish to begin using predictive

Organizations that do not use predictive analytics in their business operations are being left behind by competitors that do.

analytics in their business operations are often not sure how to begin to implement predictive-analytics-based business processes, and are understandably reluctant to invest significant resources in an uncertain venture. This article provides some direction on how to establish a predictive analytics program based on discussions with people from a number of organizations that have successfully implemented these programs in-house over the past 15 years. The following four steps summarize these successful implementations.

1. Start small and take a "learning by doing" approach.
2. Develop an initial list of possible predictive analytics projects that address frequent and important business decisions in your organization.
3. Select projects from the initial list that make use of well-known metrics for predicting outcomes.
4. Compare the results of a new predictive-analytics-based business process to the incumbent process used to make the decision.

Organizations that have had successes using this approach become firm believers in the value of predictive analytics, and as they begin to understand the possible return on investment, they devote more resources to the effort.

Start Small and Take a "Learning by Doing" Approach

Individuals involved in implementing successful in-house predictive analytics programs indicate that they started with one or two people who believed their organizations' business processes could be improved by the use of predictive analytics tools, and undertook one, or in some cases a few, predictive analytics "skunk works" projects to confirm or refute their beliefs. This early experimental approach has a number of benefits: (1) it involves an initial low financial commitment with respect to software (Excel's very limited statistical tools are a common starting point, and open source software is available and works well for many projects) and to personnel (existing personnel are used); (2) the organization is able to develop internal expertise in this area, which it can leverage in the future; (3) the organization develops a better understanding of what is and what is not possible with predictive analytics; (4) it enables the ability to assess the possible benefits from using predictive analytics to drive business processes, but in a limited way that limits the downside risk of an unsuccessful experiment; and (5) several successful small projects builds managerial confidence in the approach, enhancing organizational buy-in.

The one impediment to implementing this approach is the need to have one or two existing staff members with the willingness to take on new challenges, a basic set of computer skills, and some time to experiment with the predictive analytics tools.

The one impediment to implementing this approach is the need to have one or two existing staff members with the willingness to take on new challenges, a basic set of computer skills, and some time to experiment with the predictive analytics tools to develop an understanding of how the tools work. Advanced statistical training is a help, but is hardly a prerequisite. In fact, personnel who have a good understanding of the business, coupled with some familiarity with basic statistical concepts, will typically be more successful in the initial startup stages, than personnel with advanced statistical training but a limited understanding of the business. This is particularly important when there are important metrics related to the behavior of interest (Point 3 above) that are well known to people who work in this line of business, but may well not be known to someone coming from outside of this area. As one head of an analytics group told us, "When you question PhD statisticians about what they are doing, they want to know which of their hundred variables you don't understand." Of course, as the analytics capabilities of the organization grow, deep specialized skills in statistics, machine learning, and data management become very valuable.

Thinking through how predictive analytics tools can be used to help decision making is a crucial first step to developing a successful in-house predictive analytics program.

Develop a List of Possible Predictive Analytics Projects that Address Important Business Decisions

Thinking through how predictive analytics tools can be used to help aid business decision making is a crucial first step to developing a successful in-house predictive analytics program. A useful way to start this process is to begin with your organization's key performance indicators, or KPIs. Specifically, think through what underlying business decisions drive a particular KPI, and then think of ways that predictive analytics can be used to better inform those decisions.

To illustrate this, take the case of a marketing department for a firm that has as one of its KPIs the ROI for a direct marketing campaign targeted to existing customers. Two key business decisions that need to be made that have a direct impact on the KPI is how many customers to contact in the campaign, and which customers to contact. In this instance, a predictive model derived from the history of customer responses that provides an estimate of each

customer's probability of responding favorably if contacted in the campaign is extremely useful in informing these two decisions. The expected probability that a customer will respond favorably to the campaign, multiplied by an estimated dollar value of the response, allows the firm to calculate whether the expected return of the contact exceeds the cost of the contact. Only customers for whom the expected return of the contact exceeds the cost of the contact (or the cost of the contact plus some increment) should be included in the contact list for the campaign, thereby identifying both which customers to contact and the total number of customers to contact.

As a second example, consider a group in a credit union that is responsible for consumer personal loans. One KPI this group is likely to have is maintaining an "acceptable" default rate on these loans. The business decision that is directly related to this KPI is whether to accept or reject the personal loan application made by one of the credit union's members. A critical input to this decision that predictive analytics can provide is an estimate of the probability that a particular member will default on the loan if that member's application is approved.

In both of these cases, the analysis informs the decision, it does not make the decision. In the case of the direct marketing campaign, the model may suggest contacting far fewer customers than similar non-model based campaigns in the past. Managers may reasonably decide to be cautious and make some reduction without going all the way to the model recommended number of contacts. If the results validate the model, it can be used with more confidence in the next campaign. Marketers may also believe that simply contacting the customers, even if they don't respond, has some value, particularly with those who are near the cutoff expected return value. In the case of loan default, false negative predictions – credit union members who are actually good risks and valued customers are predicted to be bad risks by the analysis – can hurt customer goodwill. The loan officer may wish to obtain additional information to determine if the loan should be granted, or possibly make an experience-based judgment call. Both of these scenarios emphasize the critical importance of knowledge of the business when developing predictive analytics capabilities.

Select Projects from the Initial List that Make Use of Well-Known "Metrics" for Predicting Behavior

A predictive model requires a set of available metrics from in-house databases and/or third party databases that are combined to generate predictions of the behavior of interest.

While the use of predictive analytics to drive business processes may be new to your organization, they have been in use by some organizations for a very long time. As a result, there is often no need to re-invent the wheel. The ability to piggyback on what others have already discovered is likely to be particularly important for organizations just developing in-house predictive analytics programs. Your problem will likely have important predictor variables beyond the well-known ones, but the well-known ones are likely to provide a large head start in developing models.

The use of predictive analytics to drive business processes have been in use for a very long time. As a result, there is often no need to re-invent the wheel.

Returning to our direct marketing example, consider an industry that has made extensive use of predictive analytics and has long experience with this problem, the catalog retail industry. The key decision is whether to mail an existing customer a catalog or not. Three metrics that have well established predictive efficacy in this application domain are known as recency, frequency, and monetary value measures, often collectively referred to as RFM measures. Recency is often measured as the number of days since the last purchase made by a customer occurred, frequency is the number of purchase occasions of a customer (typically measured as the number of unique purchase dates) over a recent specified time period (for example, in the past year), and monetary value is the value of all purchases made by a customer over a recent specified time period. While RFM measures have their roots in direct marketing catalog retailing, they have proven valuable in other domains where the objective is to predict the behavior of existing customers.

Other verticals and/or domains have their own "go to" metrics for making predictions. In the case of loan defaults, the loan amount, loan duration, the applicant's current saving and checking account balances, and the applicant's current indebtedness are well established predictors of loan defaults. In the case of sales forecasting, seasonal indicators (e.g., month of the year), a time trend, prices of the focal product and close substitutes are often important predictors. Google searches can often be used to find relevant articles, blog posts, slide decks, and other resources that present predictive models for a particular application domain that can be used to develop a set of likely predictor

variables for a particular application. Doing some homework upfront often pays large dividends in the model building process.

Compare the Results of a New Predictive-Analytics-Based Business Process to the Incumbent Process

Testing and experimentation are an essential part of the use of predictive analytics tools. The goal of these tests is to objectively compare the performance of the predictive-analytics-based business process to the incumbent process used to make the same business decision. Favorable results in these types of tests increase managers' trust in this approach, and leads to greater levels of organizational buy-in. It can be conducted using A/B testing methods, which are becoming more common, or through retrospective testing.

A/B testing involves the use of two different samples of relevant "units" (customers, outlets, employees, or accounts, for example) with one sample of units selected on the basis of a predictive model and the other using the incumbent business process. The test results are based on comparing the performance of the two samples on the relevant KPIs (increasing incremental gross margin, favorable response rates, decreased churn rates, or decreased default rates, for example).

Retrospective testing does not require a live experiment. It uses only historical data from two time periods, where the later period has the results of the incumbent process decision. A predictive model is developed using data from an earlier historical time period which is then used to determine what decision should have been made in the later historical time period. The results of the model based decision are then compared to the actual decision made in the later time period based on the relevant KPIs. An important limitation of this approach is that often comparisons can only be made for units that were selected by the incumbent process, likely understating the potential benefit that could be obtained if the model-based process was applied to all relevant units. To make this point clearer, consider the case of a selecting which customers a retailer should include in a direct mail campaign. Only customer responses for the customers contacted by the incumbent process are available, therefore the two business processes can only be directly compared for customers that the incumbent process selected. The potential benefits that would accrue from customers that the model based process would have selected, but were not selected by the incumbent process, cannot be measured. Consequently, use of the retrospective test can have results that are artificially skewed in favor of the incumbent business process.

Testing and experimentation are an essential part of the use of predictive analytics tools. The goal of these tests is to objectively compare the performance of the predictive-analytics-based business process to the incumbent process used to make the same business decision.

The skunk works approach to implementation is essentially a proof of concept. Once it produces some successes, more individuals become enthusiastic and desire the expansion of predictive analytics capabilities.


One of the benefits of retrospective testing relative to A/B testing, since it relies solely on historical data and requires no specialized experimental procedures, is that it is less expensive to use. Because the predictive analytics process is not used in actual practice, it also entails minimal downside risk in the very unlikely event that the predictive-analytics-based business process is a complete failure.

In addition to the likelihood that retrospective tests will underestimate the value of a predictive analytics approach, they require a stable environment and consistent business practices across the time periods. For example, in deciding whether to use the method for a 2012 spring garden furniture mailer, a lawn and garden retailer could reasonably use data from its 2011 spring garden furniture mailer to retrospectively compare a predictive analytics target selection method to its incumbent method. However, the data could not be used to evaluate differences in selection processes for a proposed new fall lawn care mailer.

In general A/B testing will give better information than retrospective testing for the reasons given above. However, for organizations just beginning to experiment with predictive analytics, the much easier implementation of retrospective testing makes it a reasonable choice. Keep in mind that the approach is likely to understate the value of predictive analytics.

What may appear to be a small potential benefit from its use may in fact be a large benefit in a real application.

Scaling Up

The skunk works approach to implementation is essentially a proof of concept. Once it produces some successes, more individuals become enthusiastic and desire the expansion of predictive analytics capabilities. At this point organizations need to recruit individuals with more expertise. The ideal candidates, and the most sought after, have strong analytical skills, an understanding of business problems, and the ability to communicate with front line managers. These individuals are able to guide the development of the necessary IT infrastructure and the acquisition of better tools. Communication skills are critical, as education and persuasion of potential users throughout the company will be essential to replace incumbent processes with predictive-analytics-based business processes. Ensuring that the analytics stay closely linked to the processes facilitates adoption. The result is an organization that has better information, makes better decisions, and competes more effectively. 

About the Authors

Daniel Putler is the Data Artisan in Residence at Alteryx, the leading platform for strategic analytics. At Alteryx, Dan is responsible for developing and implementing the product

road map for predictive analytics. He has 30 years of experience in developing predictive analytics models for companies and organizations that cover a large number of industry verticals, ranging from the performing arts to B2B financial services. He is co-author of the book *Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R*, which was published in May by Chapman and Hall / CRC Press. Prior to joining Alteryx, Dan was a professor of marketing and marketing research at the University of British Columbia's Sauder School of Business and Purdue University's Krannert School of Management.

Robert Krider is Professor of Marketing at the Beedie School of Business, Simon Fraser University, British Columbia Canada. He has taught Customer Analytics to undergraduate and graduate business students at SFU, the Hong Kong University of Science and Technology, the University of Pforzheim, Germany, and the City University of Hong Kong. His academic articles have been published in the top marketing journals, and he is co-author of *Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R* (Chapman Hall / CRC Press.) Robert received his PhD in Marketing from the University of British Columbia, as well as an MSc in Geophysics and a BSc in Physics.

