CGI

Experience the commitment®

WHITE PAPER

Predictive analytics

The rise and value of predictive analytics in enterprise decision making

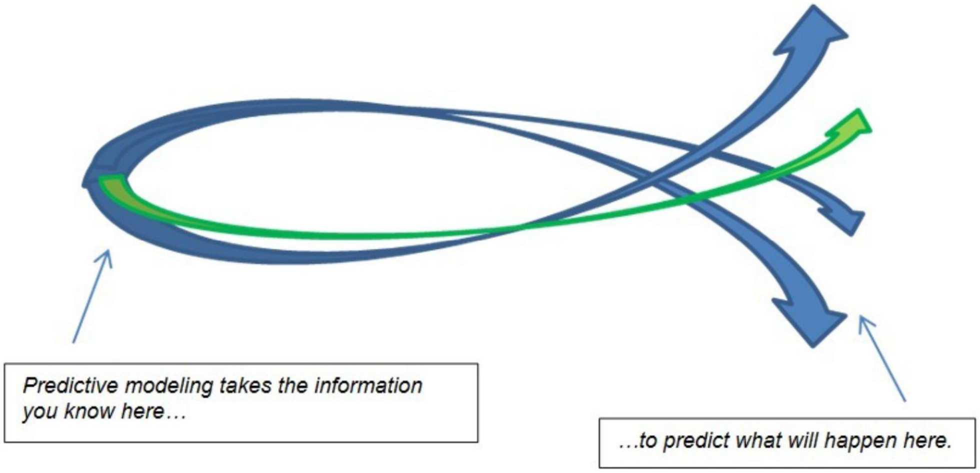
“Give me a long enough lever and a place to stand, and I can move the Earth.” Archimedes, 250 B.C.

In the past few years, predictive analytics has gone from an exotic technique practiced in just a few niches, to a competitive weapon with a rapidly expanding range of uses. The increasing adoption of predictive analytics is fueled by converging trends: the Big Data phenomenon, ever-improving tools for data analysis, and a steady stream of demonstrated successes in new applications. The modern analyst would say, “Give me enough data, and I can predict anything.”

The way predictive models produce value is simple in concept; they make it possible to make more right decisions, more quickly, and with less expense. They can provide support for human decisions, making them more efficient and effective, or in some cases, they can be used to automate an entire decision-making process.

A classic example of predictive analytics at work is credit scoring. Credit risk models, which use information from each loan application to predict the risk of taking a loss, have been built and refined over the years to the point where they now play indispensable roles in credit decisions. The consumer credit industry as we know it today could not operate without predictive credit risk models.

Credit scoring is demonstrably better than unaided human judgment in both accuracy and efficiency when applied to high volume lending situations such as credit cards. So much so, that any company in the credit industry that does *not* use it is at a significant competitive disadvantage.



Banks were early adopters, but now the range of applications and organizations using predictive analytics successfully have multiplied:

- ***Direct marketing and sales.*** Leads coming in from a company's website can be scored to determine the probability of a sale and to set the proper follow-up priority. Campaigns can be targeted to the candidates most likely to respond.
- ***Customer relationships*.** Customer characteristics and behavior are strongly predictive of attrition (e.g., mobile phone contracts and credit cards). Attrition or "churn" models help companies set strategies to reduce churn rates via communications and special offers.
- ***Pricing optimization*.** With sufficient data, the relationship between demand and price can be modeled for any product and then used to determine the best pricing strategy. Analytical pricing and revenue management are used extensively in the air travel, hospitality, consumer packaged goods and retail banking sectors and are starting to enter new domains such as toll roads and retail e-commerce.
- ***Health outcomes*.** Models connecting symptoms and treatments to outcomes are seeing wider use by providers. For example, a model can predict the likelihood that a patient presenting a certain set of symptoms is actually suffering a heart attack, helping ER staff determine treatment and urgency.
- ***Insurance fraud***. Many types of fraud have predictable patterns and can be identified using statistical models for the purpose of prevention or for after-the-fact investigation and recovery.
- ***Improper public benefits payments and fraud***. Health, welfare, unemployment, housing and other benefits are sometimes paid when they should not be, wasting taxpayers' money and making benefits less available to those who deserve them. Models similar to those used in insurance fraud help prevent and recover these losses.
- ***Tax collections.*** Likely cases of additional tax owed (due to non-filers, underreporting and inflated refunds) can be identified. The IRS and many state governments use revenue collection models and are continually improving them.
- ***Predicting and preventing street crime, domestic abuse and terrorism***. In addition to link analysis techniques for investigating crimes, predictive models help determine high-risk situations and hotspots for preventive action.

About this paper

Predictive analytics is on the rise as the number of successful applications continues to increase. Predictive models can be used to generate better decisions, greater consistency, and lower costs.

Top areas in which predictive models are generating significant value for organizations include marketing, customer retention, pricing optimization and fraud prevention—and the list continues to grow.

This paper discusses how predictive models are built, ideal situations for applying them, calculating their return on investment, key predictive modeling trends and more. With predictive analytics, organizations in both government and industry can get more value from their data, improve their decision making and gain a stronger competitive advantage.

BUILDING EFFECTIVE PREDICTIVE MODELS

Predictive models require data. Building, testing and refining these models require data that describes 1) what's known at the time a prediction needs to be made, and 2) the eventual outcome. For example, to develop a model for heart attack risk presented by patients coming into the ER, we'd need to have data describing patient symptoms when they arrived, and then the subsequent outcome (were they suffering a heart attack or not). The ability to generate data with these characteristics is a critical factor in the success of a predictive modeling application.

Statistical techniques, such as linear regression and neural networks, are then applied to identify predictors and calculate the actual models. Software from the SAS Institute, IBM's SPSS, and the open-source statistical toolset "R" are often used for this modeling analysis step.

After assembling the data, the analysts may find 20 predictive factors that are known for each patient (in our ER example) and assign weights to them using statistical software (e.g., +50 points for abnormally low blood pressure). The statistical software uses algorithms to optimize the model weighting factors, so that the combination produces the most accurate predictions possible with the available data.

The resulting "score," combining all the factors and their weights, will be an effective risk index that can be used as a decision criterion along with other rules for patient treatment. The score will be not only correlated with cardiac risk, but can be calibrated to have a specific mathematical relationship. This is a crucial advantage; it takes this risk from being an "unknown unknown" to a "known unknown" or, in other words, a *calculated* risk.

HIGH-VALUE APPLICATIONS FOR PREDICTIVE MODELING

Most business processes in most organizations have the potential to benefit from predictive modeling. That said, there are certain situations where predictive models can be especially beneficial in delivering a great deal of value:

- Processes that require a large number of similar decisions
- Where the outcomes have a significant impact, i.e., where there's a lot at stake in terms of money or lives
- Where there's abundant information in electronic data form available on which to base decisions and measure outcomes
- Where it's possible to insert a model calculation into the actual business process, either to automate decisions or to support human decision makers

Credit scoring meets the criteria for high-value application:

- There is a steady stream of loan requests accompanied by information about the borrowers
- The result of each decision has a big financial impact and is captured in loan accounting systems
- The models fit into the loan decision processes in a logical way to support human decisions

So does assessing cardiac risk:

- Patient symptoms are now captured in electronic health records, as are diagnoses and subsequent outcomes
- Decisions about treatment have life-or-death implications
- It's possible to use a model to guide immediate treatment decisions once the patient's information is entered when they come in to the ER.

Credit scoring for banks and lenders is just one of many areas where predictive analytics is driving value.

Using a model to predict a crucial business outcome, an organization can turn an "unknown unknown" into a "known unknown" or, in other words, a calculated risk.

Most business processes in most organizations have the potential to benefit from predictive modeling.

CALCULATING THE ROI OF PREDICTIVE ANALYTICS

In many cases, it's possible to measure the potential benefits and even estimate the return on investment of a predictive model using a simple method—the swap set. As shown in the table below, the swap set is the set of improved decisions made possible by a predictive model.

An example application are sales leads coming into a company's website. Once received, the leads are either assigned to one of the inside sales representatives for immediate follow up, or they receive an automated email response.

The company wants to increase sales with the same staffing level, so it develops a predictive model to measure the probability of a lead converting to a sale. The assignment rules are changed to incorporate the model by assigning high-probability leads to the sales reps. The swap set compares the results of the former assignment logic, which was based on intuitive logic, with the new model-driven logic. The result is an increase of $200,000 per month.

CGI's viewpoint is that predictive models provide an extremely effective way to get more value from data, and they have a wide range of applications that have not come even close to being fully exploited yet.

Existing processes (predictive-model supported process)	Assigned to rep	Email only	Total leads
Assigned to rep	1,000 leads	1,000 leads	2,000 leads $1,000,000 sales
Email only	1,000 leads	7,000 leads	8,000 leads $500,000 sales
Total leads	2,000 leads $700,000 sales	8,000 leads $600,000 sales	10,000 leads $1,500,000 sales (model) $1,300,000 sales (former) +$200,000 from scoring

Swap set illustration for sales lead scoring example. Use of a lead assignment model results in different treatment decisions compared to what would have been decided previously, and generates improved results with equal sales cost.

PREDICTIVE MODELING TRENDS

The biggest set of changes and advances in predictive modeling are coming about as a result of the explosion in unstructured data—text documents, video, voice and images—accompanied by rapidly improving analytical techniques. In a nutshell, predictive modeling requires structured information—the kind found in relational databases. To make unstructured data sets useful for this kind of analysis, structured information must be extracted from them first.

One example is sentiment analysis from Web posts. Information can be found in customer posts on forums, blogs and other sources that predict customer satisfaction and sales trends for new products. It would be nearly impossible, however, to try to build a predictive model directly from the text in the posts themselves. An extraction step is needed to get usable information in the form of keywords, phrases and meaning from the text in the posts, as shown in the graphic below. Then, it's possible to look for the correlation between instances of the phrase "problems with the product", for example, and spikes in customer service calls.

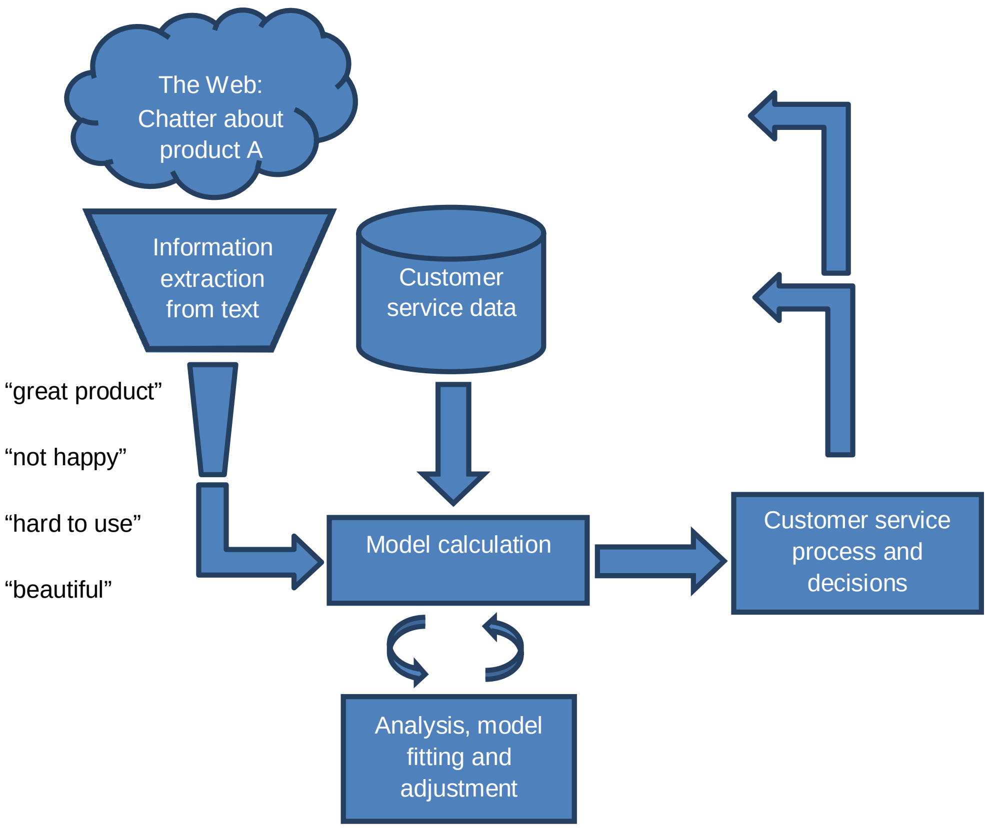


Illustration of information flow and process for a sentiment analysis application

Every form of unstructured data (e.g., text, images, video and sound) is accompanied by a set of extraction techniques that enable usable information to be pulled out of the raw data and used for analysis, including predictive modeling. With that step taken, the principles for finding high-value applications of predictive modeling are the same for unstructured as for structured information: high-value decisions, data in computerized form on both outcomes and potential predictors, and a way to insert a model into the actual decision process.

CONCLUSION

CGI’s viewpoint is that predictive models provide an extremely effective way to get more value from data, and they have a wide range of applications that have not come even close to being fully exploited yet. The Big Data phenomenon will only increase the number of high-value uses for predictive modeling in government and industry.

At the same time, predictive analytics are not appropriate or even feasible for all applications. As with any powerful technology, care must be taken to implement predictive models pragmatically in ways that produce real value.

An important challenge to greater adoption is that the talent needed to analyze data, create models and implement them successfully is in short supply. Organizations looking to get more value from the their data assets through predictive analytics will be wise to invest in analytical training and mentoring programs, along with the expertise of partner firms.