Real world predictive analytics

Putting analysis into action for visible results

Contents:

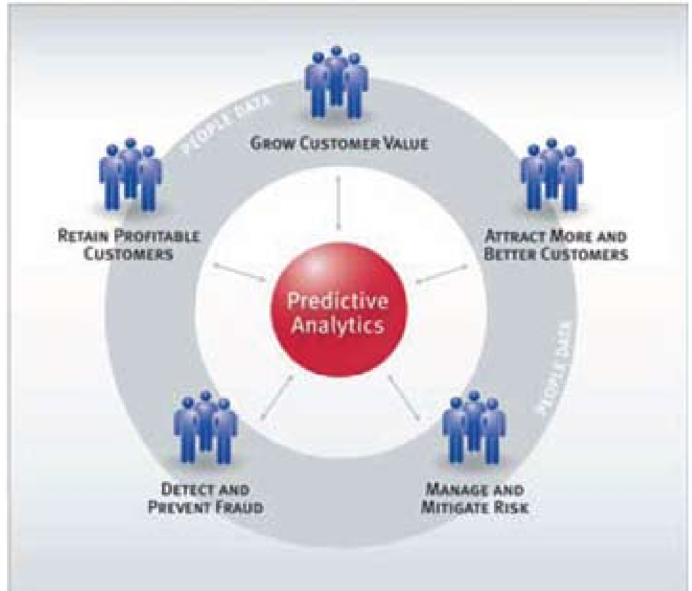
Introduction

Predictive analytics connects data to effective action by drawing reliable conclusions about current conditions and future events. This approach, rooted in advanced mathematical analysis, can seem esoteric and may almost be perceived as an academic discipline. Organizations may read case studies that list high-ROI applications of predictive analytics but often struggle to understand, in ways relevant to their business, how predictive analytics can be applied in numerous operational areas to improve outcomes significantly. By discussing a number of typical scenarios in which predictive analytics addresses key business issues and illustrating ways analytical capabilities are deployed in business processes and the operational systems that support them, this white paper aims to show how predictive analytics delivers business value across the enterprise.

Predictive analytics in business

In predictive analytics, advanced analytical algorithms process historical data, "learn" what has happened in the past and create models that can be applied to make judgments about current or future cases. In theory, this approach can be applied wherever historical data on past outcomes exists (such as records of which customers responded or didn't respond to a particular offers, or which borrowers defaulted on loans). In practice, it is most typically used to address five business objectives, which all center on the idea of *gaining value from customer relationships*:

- To **attract** more (and better) customers
- To **grow** the value of existing customers, through cross-selling or up-selling products or services
- To **retain** the most valuable customers
- To manage and mitigate the **risk** inherent in transacting with customers
- To detect and prevent **fraud**



Key business objectives, interconnected through predictive analytics

Applying predictive analytics in any one of these areas can generate significant value. Organizations more advanced in their usage of the technology – who are able both to integrate with operational architectures and to modify business processes to take advantage of predictive capabilities – can use predictive analytics to interconnect these business areas in ways that boost returns and address the overall goal of *profitable* revenue growth.

For example:

- Retention analysis may reveal which product cross-sales increase customer loyalty
- Growth analysis gives the profile of customers with the greatest value potential, which can be used to try to attract and acquire similar customers
- Fraud analysis exposes the characteristics of customers who should be avoided in order to minimize the risk of fraud-related losses

Predictive analysis enables companies to predict which high net worth customers are most likely to respond to a marketing offer or defect to a competitor.

Components of a solution

To understand how predictive analytics delivers value, it is useful to think about a predictive analytics solution as comprising three main elements:

Data-driven business decisions

At the top level, predictive analytics enables better business decisions. While analysis may reveal new insights that can drive far-reaching, strategic decisions by senior management and deliver step changes in business value, it is more common to see these insights applied at the level of individual cases – enhancing key business decisions that are made frequently and repeatedly, where improvement leads to a higher proportion of good outcomes and a clearly measurable, incremental ROI. It can be helpful to think of these decisions as "*if we could make better decisions about <X>, we could deliver greater value by <Y>*."
For example:

- If we could reliably predict which of our high net worth customers were likely to defect to a competitor, we could intervene and offer incentives to persuade them to stay loyal.
- If we knew how likely it was that each of our customers would respond to a particular cross-sell offer, we could reduce the size (and cost) of campaigns by not targeting people unlikely to respond; increase response rates and revenues; and avoid the situation in which customers who have no interest in the offer see it as junk mail.
- If we could accurately assess the risk of each insurance claim as we receive it, we could fast-track safe claims to reduce costs and increase customer satisfaction and loyalty, while increasing our fraud detection rate by ensuring that our investigative resources focus on genuinely high-risk cases.

*"**Deployment**" refers to the implementation of the analysis results into business processes or operational systems.*

Deployment into processes and systems

In predictive analytics, "deployment" describes the integration of the results of analysis – for example, pre-calculated propensity scores or a "live" predictive model – with business processes and the operational systems that support them to make better decisions and drive better outcomes.

In some solutions, this may be simple – one point of integration in one process and system. For example:

- At a single point in the processing of tax returns, a predictive model scores every return on the likelihood of non-compliance, and adds those with high-scores to a list to be audited by an investigations team.

In others, it may be more complex:

- In insurance risk management, models are integrated with the claims handling process at the point of claim creation to assess risk and recommend a routing (fast-track/investigate/normal) for each claim. For claims following the normal process, the models could also be run at later points in the process to re-assess risk (and re-route claims as appropriate) as new data becomes available. At all these points, the model invocation is integrated with the claims handling system that supports the process.
- A model for up-selling may have a single, simple purpose – to recommend whether a specific product should be offered to a specific customer – but may be integrated with multiple processes and systems to ensure offers are made through the most appropriate channels and touchpoints for each customer. In a bank, such a model could be used in several ways, such as:
  - In the direct marketing process, to select customers to receive outbound mailing or calls, integrated with a campaign management system
  - Within the online banking system, making suggestions to customers who log on to service their account
  - Coupled to the transaction handling processes, making offers when customers visit a branch or use an ATM

*The **analytical process** forms the foundation for a business solution using a three-step process: **Capture** the data, **predict** the outcomes and **act** on the results.*

The analytical process

Behind the scenes, the foundation of the business solution is the analytical process that produces the results to be delivered to the front line. While this may be as complex as any given situation requires, the analytical process follows three basic steps:

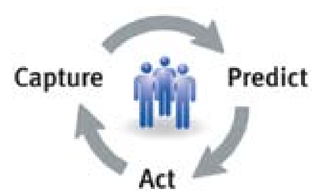
- **Capture** a holistic data view of the customer (referred to as the "analytical data view"). This will normally consist of up to four types of data:
  - *Descriptive* data: usually a mix of self-declared information and externally-sourced (geo)demographics
  - *Behavioral* data: this could be as simple as transaction records – who bought what and when – or details of how customers use a product or service. For a wireless phone operator, it could include detailed information on calling patterns. For a credit card provider, it could include details of how customers spend with their card and how they pay off their bill.
  - *Interaction* data: the details of how customers interact with the company through its various channels. This might include details of web site visits by registered customers (though mapped from "click-level" data to business-meaningful "events" that happen during visits), or "unstructured" text emails to customer service contacts, or transcripts of call center conversations. Also in this category is data originating from customers' interactions with each other. In the era of Web 2.0,

this includes discussions on forums and blog postings, as well as "social networks" representing how customers communicate with and influence others. Integrating such data into the analytical data view requires specialized technologies, such as text mining, web behavior analysis and social network analysis.
- *Attitudinal* data: people's needs, preferences, opinions and desires. In the past, these have often been collected in the course of surveys conducted for market research or to assess customer satisfaction. However, in these cases they have been used only in an anonymized, aggregated form. Today, many organizations recognize the value of leveraging this data at the individual customer level, linked to the other elements of the analytical data view.

Predict outcomes in current or future cases by analyzing the data, using advanced algorithms, to create predictive models. A prediction is the "raw" output of a model. It may be the propensity for a customer to behave in a particular way (e.g., the likelihood that a customer who owes money will pay it back), or a selection of one of several options (e.g., which of the available collection strategies will be most effective in persuading the debtor to pay).

Act on the results of analysis. This involves:
- "Decision management": combining the results of possibly multiple models with business logic (rules, policies, exclusions etc.) to arbitrate between different possible actions and decide on the right one to be taken
- Integrating these decisions and actions into key points in the relevant business processes and the operational systems that support them

"**Capture, Predict, Act**" takes us from collected data, through advanced analysis, to the successful deployment of analytical results to improve business processes. It is worth noting, though, this is actually a "virtuous cycle," producing continuous improvement: new data captured at the point of action (i.e., during customer interactions) enhances the analytical data view, enabling more accurate predictions and driving better decisions with a greater proportion of positive outcomes and, hence, higher returns.



In many cases, elements of the analytical process will be automated to enable this iterative approach and to address other requirements such as monitoring of model performance and scalability to multiple related application areas.

Examples: Predictive analytics has helped three real-world organizations solve problems and improve their operations.

Example scenarios

Following are three real-world examples that demonstrate different aspects of applying predictive analytics to business problems, and the range of areas to which this technology can be applied. The three scenarios described here are:

- Identifying and retaining potential defectors in a wireless telecommunications company
- Growing business profitably by combining cross-sell models with credit risk models in personal finance
- Identifying and targeting risks at border crossings

The first two scenarios apply to traditional customer situations; the final scenario shows predictive analytics applied to citizens rather than customers, though the approach, principles and practices remain the same.

Scenario 1: Identifying and retaining potential defectors (telecommunications)

The data-driven business decision:
If we knew which of our on-contract customers were likely to defect to a competitor, and which of these were worth keeping, we could intervene in a timely manner and take steps to keep them loyal. This would reduce our churn rate, ensuring an ongoing revenue stream from high-value customers, reducing acquisition costs for replacing defectors and maintain/increase our market share.

Deployment into processes and systems:
With such a critical issue, the organization has to take every possible opportunity to retain at-risk, high-value customers. This means the predictive models have to be available to identify such customers at every point of interaction. This will likely involve:

- Outbound retention offers to probable defectors by mail, email and SMS
- Outbound calling campaigns to the most valuable customers who are at most critical risk
- Potential churners flagged, and actions taken, during inbound interactions:
  - Recommendations made to retail store staff when customers visit
  - Service center agents prompted to make offers during customer calls
  - Offers popped to customers logging on to the provider's website

Acceptance or declination of offers will be tracked and reported on along with attributable improvements in churn rate.

The analytical process:
Telecommunications companies hold rich data on their customers. Call detail records (CDRs) can be used as the basis of extremely rich behavioral descriptions, including call volumes, peak vs. off peak usage, "social network" patterns of calling, roaming behavior and many other factors. Descriptive information may be limited to what was declared at the time the contract was taken out, or may be augmented with purchased geo-demographic information based on postal/ZIP code.

Interaction data in the form of text notes from service center conversations may indicate retention "red flags" in terms of complaints raised, while usage of certain Web facilities (e.g., looking up contract terms and conditions) may also indicate that customers are considering switching suppliers. Attitudinal data may be key both for knowing what has made a potential churners dissatisfied and what type of offer or treatment may motivate them to stay loyal.

The core model to be built is one that scores a customer on how likely it is that he or she will cancel their contract within a given timescale – perhaps two-three months, a horizon which gives time for action to be taken to try to retain them. Targeting all at-risk customers, though, is expensive and may be counter-productive; only customers who are worth keeping should be targeted with retention efforts. This means that a measure of value has to be taken into account. This can either be a current measure of profitability; a traditional, static formula for calculating likely lifetime value (LTV); or a predictive model which estimates LTV dynamically for any customer. Additionally, one or more models may be used to select the retention offer most likely to persuade a given customer to stay loyal, the most effective way to approach them, the best channel through which to reach them etc.

Action may be taken by scoring customers in sync with the monthly billing cycle and associated data refresh, on both their churn propensity and value. These scores and selections, along with model-selected treatments and under the control of rules governing marketing policies, are fed to campaign management systems which run outbound, proactive retention campaigns. (The offers themselves may be formulated by marketing staff using profiles and other insights identified during the modeling process.)

However, even more effective action can be taken if the provider can be reactive to new information that indicates a sudden change in a customer's defection risk level. To do this, live models need to be deployed to inbound channels such as the service center. A customer calling to make a particular complaint may, when this fact is combined with the other data the provider holds on the customer, show a suddenly raised level of imminent defection risk. Running in real time, the retention model can recognize this and the best offer can be selected and pushed to the service agent.

Wireless telecommunications companies often operate in highly dynamic markets; customer preferences change quickly based on fashions, trends, competitors' actions and the products, services and packages the provider itself offers. Models get out of date quickly and need to be "refreshed" on new data representing the latest customer behavior and preferences. This means it will be necessary to automate model monitoring and refresh, automatically bringing the models up to date when their performance declines. Alternatively, experience may quickly lead to a strategy of regular model re-building – the models, for example, being automatically regenerated each month before the customer base is scored.

Scenario 2: Combining cross-selling with credit risk assessment (personal finance)

The data-driven business decision:
Recent financial events have demonstrated that not all business won is good business. Going forward, all our business decisions must be underpinned by an awareness of risk. In the past, it was common practice to market to customers based on their likelihood to purchase a new product. Today, we have to ensure that we don't offer products to people who will take them but who are unlikely to be able to pay for them. If we can filter our marketing offers in this way, we can ensure that the business these offers drive is truly profitable and does not increase our risk exposure.

Deployment into processes and systems:
In this case, deployment is relatively simple. The bank uses an outbound marketing approach, using a campaign management system to deliver offers to customers via direct mail and track campaign effectiveness. All that is required is for the customer lists to be selected based on propensity, then filtered to remove prospects who are risky from a credit perspective before being passed to the campaign management system for execution. Saving the purchase propensity score and the credit risk score to the customer data warehouse also makes it possible to use the combined scores to subsequently control offers of this product through other channels – for example, when customers log on to online banking or use an ATM.

The analytical process:
The two types of model – purchase propensity and credit risk – may use different but overlapping subsets of the analytical view. The propensity models will be built from historical campaign records of who responded (and who did not) to offers for similar or identical products. The credit risk models may use externally sourced risk scores as one of their inputs but will typically refine those "one size fits all" scores by adding many other variables and looking at the history of which customers proved to be good or bad risks *in this business*.

Action here will focus on how to combine the models' scores. In this example of decision management, human expertise will be added to decide how to resolve combined scores that work in opposite directions. Thresholds will be set that veto all offers when the associated risk exceeds a specified limit. Lower bands of risk, however, may be accepted under certain conditions – for example, if the propensity score is high, the predicted profit (possibly from another model) is high and the predicted *amount* of risk exposure (again, possibly from a model) is below a specified limit.

In this case, automation is required for the ongoing monitoring of the performance of both models, alerting human experts (and possibly automatically refreshing the models) as performance degrades. In particular, a lesson learned from the credit crunch is that risk models need to be kept up to date as patterns of defaulting change.

Scenario 3: Stopping suspect cars at border crossings

The data-driven business decision:
We cannot stop and search every car that crosses the border. However, if we can accurately predict the risk level associated with each vehicle – whether it may be carrying contraband, drugs, money, weapons or illegal immigrants – we can make optimal use of our inspection staff, increase our detection rates, better protect the country and its citizens and improve the experience of innocent travelers by expediting their crossing.

Deployment into processes and systems:
At each border crossing, cameras record and recognize the registration plate of every vehicle. Once this is read, directives are given on the screens in the crossing control booth – telling the supervisors to either let the vehicle pass or direct it to the secondary inspection area. If a vehicle is selected for secondary inspection, information is sent to the PDA of the inspector responsible; this shows the likelihood of each risk type (drugs, weapons, etc.), providing the inspector with guidance on what to look for. Vehicle selections, risk assessments and inspection outcomes are recorded to enable ongoing reporting on inspection rates by risk type, hit rates and false positive rates and seizure amounts and values.

The analytical process:
Models are built from the outcomes of historical inspections. The data used is keyed by vehicle registration and would include vehicle type (and hence dimensions, capacities, etc.); vehicle ownership (and hence any available information on the owner or driver); and the vehicle's history of border crossings at this or other checkpoints. Other factors giving information about the crossing would also be incorporated – for example the day of the week, time of day and prevailing weather conditions.

One model is created for each type of risk. In turning these individual models' scores into actions, scores will be combined with rules representing the best human knowledge on border risk assessment. Some of these rules will help govern how "tightly" vehicles are selected for inspection (i.e., what risk scores trigger a secondary inspection) and may be varied by pre-defined knowledge of peak traffic times, or manually in response to abnormally heavy traffic and the corresponding need to let it flow without creating a backlog of secondary inspections.

In this scenario, automation is necessary to allow tailoring of the risk solution to each border crossing. The agency controls approximately 300 crossings, and while many will show similar patterns of violation, each will have its own individual profile of what risks are likely to occur and which violations will be attempted in which ways. Ideally, each crossing point should have its own model for each risk type, built from local data. Creating and managing this number of models manually would be hugely labor-intensive and expensive, and for these reasons probably impractical. If, however, analytical experts create not the individual models but the process for building and applying the models, this process can be automated and applied, on local data, for each crossing, ensuring a "best fit" local solution that embodies national best practice.

Although this scenario is a very specific security application, it closely resembles many other deployments of predictive analytics for risk in both the public and private sectors. The basic approach of identifying the few high-risk cases among masses of "safe" cases ensures that investigative resources are used efficiently and effectively. It is also easily applied to banks intercepting money laundering transactions, insurance companies identifying potentially fraudulent claims, revenue authorities targeting tax evaders and to many other areas.

Three key points common to all scenarios presented here are:

- Accurate analytics requires the convergence of analytics data, business processes and architecture
- The holistic analytical view should be treated as an aspiration not a practical requirement
- Tracking the outcomes of decisions driven by predictive models helps organizations continuously improve

Becoming a Predictive Enterprise is a journey that requires a roadmap to achieve optimal business value.

Key components for success in predictive analytics

What can we observe from these scenarios that is crucial for the success of predictive analytics solutions? Three key points are common to all the scenarios described:

1. Analytics cannot be considered or applied in a vacuum. To be effective, organizations need to take an approach that focuses on the **convergence** of analytics, business processes and architecture. Predictive analytics delivers value to an organization when used to improve decision making (and hence outcomes) in key business processes and, therefore, needs to be integrated with the architecture and operational systems that support these processes.

Critical Convergence

2. Data is crucial. It is impractical to attempt to capture all possible information on customers (or cases) before beginning analysis. Rather, the **holistic analytical data view** should be treated as an aspiration; organizations that continually enrich the data available for analysis will see incremental improvements in analytical effectiveness and in ROI.
3. Predictive analytics provides the means for an organization to learn and improve continuously. **Closing the loop** – tracking the outcomes of decisions driven by predictive models and capturing data at every interaction – guarantees that the quality and consistency of critical business decisions improves over time and that effective decisions continue to be made as situations change and evolve.

Getting there

A Predictive Enterprise is able to apply advanced analytics effectively across its organization, as in the example scenarios described here. No organization, however, becomes a Predictive Enterprise overnight. Rather, the most effective approach is to treat the adoption of predictive analytics as a journey. Creating a "roadmap" for this journey involves identifying possible applications of predictive analytics within the business and prioritizing them based on business value and analytical fit. In this way, organizations can incrementally transform their operations, enabling improvements and higher returns in more and more areas across the enterprise – continuing to do what they do, but doing it better as data-driven decision making becomes prevalent throughout their business processes.

About IBM Business Analytics

IBM Business Analytics software delivers complete, consistent and accurate information that decision-makers trust to improve business performance. A comprehensive portfolio of business intelligence, predictive analytics, financial performance and strategy management, and analytic applications provides clear, immediate and actionable insights into current performance and the ability to predict future outcomes. Combined with rich industry solutions, proven practices and professional services, organizations of every size can drive the highest productivity, confidently automate decisions and deliver better results.

As part of this portfolio, IBM SPSS Predictive Analytics software helps organizations predict future events and proactively act upon that insight to drive better business outcomes. Commercial, government and academic customers worldwide rely on IBM SPSS technology as a competitive advantage in attracting, retaining and growing customers, while reducing fraud and mitigating risk. By incorporating IBM SPSS software into their daily operations, organizations become predictive enterprises – able to direct and automate decisions to meet business goals and achieve measurable competitive advantage. For further information or to reach a representative visit www.ibm.com/spss.

IBM Corporation
Route 100
Somers, NY 10589

YTW03112-USEN-00