Research Brief

Prescriptive Analytics:
Just What the Doctor Ordered

September 2014

Written By:
Fiona McNeill
Key Takeaways

1. Being an analytically driven organization means basing decisions on data rather than on experience or intuition; one of the strengths of advanced analytics is to help see beyond current activities, into future scenarios. But knowing what is going to happen, and knowing what to do are two different things. Prescriptive analytics answers the question of what to do, providing decision option(s) based on the predicted future scenarios.

2. Text data extends predictive insights by not only bringing greater contextual awareness to understanding situations but also can define what prescriptive actions to take. And while including text insights in predictive models can improve model lift, it can also inform the scope of questions to be asked of the data and may also change what numeric/structured data factors are important to begin with.

3. Decision management systems are designed to bring together all the necessary elements needed to explicitly codify operational decisions, and when the volume and throughput of the big data is on the order of thousands and millions of events per second, conditions such data can still be assessed for analytically sound actions. In event stream processing, continuous evaluations of streaming data occur in-memory and in-stream, effectively prescribing the best action to take – in real-time.

Introduction

Triage, as a first line of defense, is used to define priorities when resources are insufficient to immediately treat all issues. As an example of business triage, in 2001, Doug Laney of METAGroup (now Gartner) first wrote about three key elements reflective of e-commerce escalation. Among those three, he recognized that the e-commerce compute infrastructures were insufficient and, as a result, was pushing enterprises to fundamentally reconsider how data was managed. Later popularized as big data, the symptoms of increasing data volumes, variety and velocity, became the triage yardsticks, to help organizations assess the degree to which they had big data, and once established, what the organizational prognosis was.

When Google and Yahoo self-diagnosed their big data challenges, they created MapReduce and Hadoop, respectively. Essentially these methods divide the data and spread the processing

\[1\] Laney, D. Application Delivery Strategies: 3D Data Management: Controlling Data Volume, Variety and Velocity, File 949, Feb. 6, 2001

\[2\] Vance, A. Hadoop, a Free Software Program, Finds Uses Beyond Search, New York Time, March 16, 2009
across thousands of low commodity computers, operating in parallel. In this new data processing archetype, each disk produces interim results which are then collected to form output based on a pre-defined format. This technology is quickly becoming mainstream, with Hadoop in 65 percent of big data proofs of concept in North America and about 10 percent in Europe, rising to more than half of the organizations surveyed by 2016.

Great for indexing enormous quantities of both structured and unstructured data, as well as ensuring high availability for ad-hoc queries, these new methods to store and retrieve big data allowed new questions to be asked simply because all the data was accessible. However, going beyond queries to advanced analytics required more efficient compute methods.

Analytic algorithms in this framework needed to make repeated calls to the disk during processing, reading and writing each time. In-memory computing resolved this, storing large blocks of data directly in the random access memory (RAM) of these distributed commodity server clusters, and keeping the data in-memory for continued and complex analysis. And in best in class organizations, roughly two in five (38 percent) had implemented this technology in 2013, being early adopters and deriving benefit.

With in-memory computing opening the big data analytic gates, both the type and scope of business problems that can be examined is radically changing. Solving historic mysteries of humankind in genetics, predicting future events in human health, and improving existing processes in business, big data analytics is clearly a new era of understanding.

**Predictive and Prescriptive Analytics**

Being an analytically driven organization means basing decisions on data rather than on experience or intuition. And according to the Analytics Mandate, a global executive survey

---

with over 2,000 respondents, over 85 percent of businesses agree that, at least somewhat, they need to increase the use of data analytics in order to drive decisions. And yet, as more organizations recognize the competitive advantage of being analytical, that advantage wanes as competitors build this same capability. So while the Analytics Mandate describes the behaviors, values and outcomes associated with developing an analytics culture, it also recognizes that using similar data to develop similar models doesn’t provide unique benefits.

One of the strengths of advanced analytics is to help see beyond current activities, into future scenarios, by predicting (with a known certainty) the potential outcome of an event. Will customers leave in the future? Who is the best candidate to hire? Are consumers likely to recommend a service or product in the future? When is it likely that there will be an energy outage? Captured as probability scores, such measures describe the likelihood that the event of interest will happen, going beyond the current facts and figures, providing an analytical crystal ball. And by knowing what is likely to happen, current decisions can be made to help drive more desired results in the future. For example, knowing that a customer is more likely to leave than stay, (say, based on decreasing purchasing patterns and increased calls into the support line), a personal call from a customer care representative or an incentive offer may help salvage the relationship. And while an experienced analyst may be able to see the emerging trend in the data, and believing that this customer is on the verge of leaving, they might have a hard time quantifying this event. How likely exactly is the customer to leave? A predictive model provides a quantitative measure to base decisions on (e.g. the model can there is an 80 percent probability that the customer will leave), without human, subjective bias.

Predictive algorithms are formulas that describe a specific scenario using historic knowledge to increase awareness of what comes next. But knowing what is going to happen, and what needs to be done about it, are two different things. That’s where prescriptive analytics comes in. Prescriptive analytics answers the question of what to do, providing decision option(s) based on the predicted future scenario.

Seldom (if ever) do events happen in isolation. It’s through the interconnections across entities, activities and time, assessing the dependencies and relationships, that we develop an understanding of what needs to be done to change future trajectories. The richness of this understanding, in turn, also determines the usefulness of the predictive models. Just as the best medicine is prescribed after a thorough examination of patient history and existing symptoms, etc. so too are the best predictions, founded on well understood scenario context. While different medications can react with each other and one medicine may work best in one

---

11 Many texts are available on this subject, one being V. Ranadive The Power to Predict: How Real Time Businesses Anticipate Customer Needs, Create Opportunities, and Beat the Competition, McGraw-Hill, NY 2006
scenario than another, under different circumstances you would have different predictions. When conditions change, predictions will also change\textsuperscript{13}.

Well-understood scenarios are fed by data that richly describe the context. The more data you have to draw from to examine the dependencies and relationships that impact the event being predicted, the better the prediction will likely be. Big data is more data, higher resolution data, and richer data. It offers details not historically available that explain the conditions under which events happen, or in other words, the context of events, activities and behaviors. Big data analytics allows us, like never before, to assess context – from a variety of data, from different sources and with less latency. Big data analytics also gives us the power to write better business prescriptions.

The ‘t-e-x-t’ in Context

Even though big data provides the opportunity to better describe predictive scenarios, if different organizations largely collect the same big data to predict churn, drive new sales, etc. the competitive advantage of utilizing big data analytics will, over time, decline, just as happened with traditional data analysis. In traditional predictive analysis, structured data, neatly defined in rows and columns, gives us answers to what will change and when. It doesn’t typically tell us how it will change or why it might change. Furthermore, the stark reality that roughly 80 percent of all business data is unstructured\textsuperscript{14} means that by using only structured data – predictions are based on only 20 percent of the scenario context. In order to change these results, we need to reframe how we define the business question.

In focusing just on the text portion of unstructured data\textsuperscript{15}, different questions can be asked when this free-form source is included. If you’ve ever filled out a survey, you know that it can be hard to quantify what you may think. It’s only in the ‘other’ box that you have the freedom to describe what’s really on your mind. Those descriptions, (i.e. the unstructured text), provide unique insight into the psyche of the respondent in often detailed, descriptive narratives. And it is in that very narrative, coming from service notes recorded in transaction systems, or online chat exchanges that understanding of how a problem occurred or why the resolution was

\textsuperscript{13} This is precisely why it’s necessary to monitor the health of predictive models. If you don’t like what is predicted to happen, based on the historic scenario, you do something different – you change the conditions, hoping that will change the outcome. And once you’ve done something different, say provide that incentive to entice the customer not to defect, you may change the scenario – and they might stay. But your model is predicting if they will defect, and yet they stayed, or you tried your best offer, and they still left. The actions taken not only create new history, they may change the behavior the model is trying to predict. As such, conditions have changed and the model context no longer reflects the right scenario for that customer. En masse, and over time, conditions will change and as such, models tend degrade, requiring updates to ensure accuracy to the current context.

\textsuperscript{14} http://en.wikipedia.org/wiki/Unstructured_data

\textsuperscript{15} Unstructured data definitions can also include audio and video data, in addition to text data
insufficient can be achieved. Text data typically provides that richer understanding, putting contextual flesh on the scenario picture.

Inclusion of text data in analysis not only provides context to structured\textsuperscript{16} data, it means you are extending beyond only using that 20 percent of available business information. In turn, this may often correspond to new sources of data, like emails, social forums, claims notes, etc. adding previously unused resources to your analytical pool. By using new sources of information as well as types of data you begin to take on a more global understanding of situational events, thereby allowing the business question to be reframed. ‘Which customers are more likely to leave?’ can become ‘What issues are potential churners likely to discuss before they leave?’ or ‘How much stock do I need for next quarter?’ might morph into ‘What distribution issues will affect our ability to restock to predicted store level requirements next quarter?’

Business questions that are posed when the context of the scenario is extended can be much more inclusive of how an organization operates. In other words, when you include text data into analysis, you can ask more in-depth questions of how and why events, activities and behaviors occur simply because you include inputs that go beyond the numbers\textsuperscript{17}. The more in-depth the analytical investigations are, the more unique they become to your organization, examining how it runs, interacts with customers/constituents and what dependencies there are in your business activities. And given they are more unique, it is more difficult for competitors to replicate the same answers, helping provide that desirable competitive advantage.

Prescriptive answers can also be found in text-based data particularly with self-reported information (e.g. social media, call center transcripts, surveys). The descriptive narrative will often detail what perceived change is needed for the author to have a different experience. For example:

- By analyzing the text data in their surveys, Alberta Parks learned that it was important to address noise issues such as complaints of noise during quieter hours because they contributed to perceptions of personal safety\textsuperscript{18}.  
- Changes for more informed web content made by one insurance provider were based on discovered themes from call center inquiries, reducing the calls per claim by 10 percent. And an online business printing company cited: “Your Web presentation of your card should be as close as possible to the printed card. Often the lettering on the card is smaller than the presentation to the point that not a few times the printing was

\textsuperscript{16} Structured data is also loosely known as numeric data
\textsuperscript{17} Now, the insights from text data might be buried in document collections - especially when there is a lot of it, but that is what text analytics is designed to decipher.
too small to ... The recommendation was to represent the ordered business in its actual size on the website – makes sense.

Another value from analyzing text data is to automatically put the unstructured data into a more traditional, structured format so that it can be used in pre-existing systems and applications. By doing so, what was previously unusable by other applications becomes useful. Structured text becomes new fields for inclusion in traditional repositories, new metadata for improving applications or as variables to be explored in business analytics tools.

Case Study: Understanding Secondary Adverse Events

- Background: In the department of Orthopedic Surgery at Denmark’s Lillebaelt Hospital, only one percent of patient records in the Department of Orthopedic Surgery at Vejle Hospital were manually reviewed. This labor-intensive task of reading the sample patient records was done four times a year by both a surgeon and a secretary – for quality control purposes. In those samples, about 30 percent of the records were found in error. This means that too few or faulty diagnoses or treatments were registered in the patient record. For example, if a patient admitted for a thigh-bone fracture caught pneumonia during hospitalization, the secondary illness and the associated pneumonia treatment weren’t registered.

- The department has since implemented the ‘Clinically Correct Time-True Registration’ system, which was designed to automatically extract diagnostic codes from the unstructured text in the surgeon’s dictation and compare to the actual recorded diagnosis codes. Customizations included coding to ensure that both Danish and Latin terms, abbreviations, etc. of this often highly technical jargon was interpreted correctly. The system has been applied in at least three different surgery departments, with an improvement potential, conservatively estimated, at over four million dollars.

- In addition to saving money, time and helping doctors spend more time with patients, all patient records can now be analyzed (vs. just the one percent sample). A new database has also been established based on the structured record data for clinical research.

Defined as new variables by text analysis, these context rich representations can also be included in advanced predictive analysis. In this way, you can statistically evaluate the degree to which text-based insights actually improve model accuracy. In other words, ask the question, ‘Does including more context actually affect predictive power?’ And in many instances, across

---

industries, it not only does but text variables have been seen to increase the significance of numeric variables that were previously not significant prior to text inclusion\textsuperscript{20}.

Including prescriptive measures from text analysis in forward thinking, predictive analysis can inform the scope of questions that can be asked and what decisions need to be made. It can also change what factors are recognized as influencing decisions in the first place.

\section*{Regulating for Real-Time}

One of the promises of using big data is that business will not only be better, but also faster. Let’s consider for a moment what that means to organizational operations. At a high level, being real-time means that incoming data can be processed as it arrives and typically without buffering delays, and processing time is typically measured in seconds and fractions of a second. Translating this to operations might mean answering credit approval requests within a second of receipt of required information while the customer is still primed to borrow from you, or serving real-time advertisements to mobile users as they surf websites. To consistently provide precise decisions in real-time, as in both of these simple examples, you need to involve some type of automation.

In operations, real-time, automated decisions also mean that the outcome that happens is the outcome that is expected to happen. Based on the credit request application, the loan that is provided is within the policy of the organization. Furthermore, under the same conditions, the realized outcome is consistently the same, desired action. Another person with the exact same loan application details would be given the same amount of credit. However, if the scenario changes, then operationally there must be some type of agility built-in so that precise decisions continue to occur, even under changing conditions.

Automation ensures that the right operational decision can consistently be executed in real-time for the same scenario each time it occurs. When scaling this to the hundredth, thousandth and millionth time, the incremental cost of making that specific decision naturally decreases. And because it’s automated, the elements that define what decision should be made are explicitly programmed. The inputs are unambiguous, the calculations are obvious and the output is defined. This not only fosters continuous model improvement, but perhaps more importantly, reduces the risk of the wrong operational decision being made and helps govern the action is within the specified policies of the organization, industry and governing regulations.

To employ good, automated decisions in operational business processes they need to be specified in a centralized system, one that is controlled with appropriate authorization rights, data access permissions, and traces each analyst’s change in the specification and testing of different scenarios. And while it can take time to codify the business rules that govern scenario definitions, time to identify the right operational data inputs, and time to select the appropriate predictive model that estimates future conditions; the security, confidence and quality assurance that occurs each time that frequent, repeated, operational decision happens tends to substantially outweigh the development effort.

Case Study: Optimizing Credit Risk Decisions

- Background: In this European financial services company, the use of analytics and predictive models to analyze customer portfolio risk was commonplace. When the portfolio fell below acceptable risk levels, a frontline employee was notified to contact the customer and adjust the portfolio accordingly. The problem was that it took too long to move the tested analytical models into the production environment where they were then registered to the employee notification system. In fact, to move the models into production required custom development that existed outside of IT – so in addition to production delays, any change because of new decision policies led to a slow time to market for the company.

- The company implemented a decision management solution in about two weeks and the analytical model deployment time was reduced by about 75 percent, helping the organization realize the desired results sooner. Reflecting the improved collaboration realized between the data scientists building the models and the business analysts optimizing the business results, new factors driving better operational processes were identified. Predictive model performance also improved because personnel had more time to spend on tuning and updating the algorithms, instead of spending their time deploying predictive models. Moreover, credit risk managers became more agile to new business conditions given they had the authorization to adjust production models themselves.

- Decision management systems are designed to bring together all the necessary elements needed to explicitly codify operational decisions, so that they are automated and executed in real-time, or in any other desired schedule. Decision management systems are used to examine scenarios, conditions, predictive models and operational data all together; elements are identified and tested for alternate business outcomes. The outcome selected becomes a prescriptive action; instructing dependent systems and/or personnel what action they are to take, just like the notification to frontline agents to call customers whose portfolio risk is calculated outside of acceptable thresholds.
Case Study: Delivering Personalized Mobile Marketing Offers

- **Background:** As a mobile marketing company, ZapFi helps retail businesses deliver real-time, personalized advertising through its network of Wi-Fi hotspots. Having the advantage of reaching consumers at the point of purchase, this direct selling method is based on sending appropriate promotions to individual consumers based on direct knowledge of current interests and characteristics.

- **The ZapFi Zones are hot spots sponsored by ZapFi customers, being merchants like banks, insurers or retailers.** Signing on to this free network, consumers provide demographics, like name, age, gender and address and their browsing behavior is captured while they are in-store, surfing on the zone. Each time they connect, their interest domain profile continues to grow, providing more details related to that specific consumers shopping behavior. Complementing the known interests (that have been identified using text analysis of the browsed web pages), with the products the retail merchant has on special promotion, a real-time match is determined with decision management technology – ultimately sending specific marketing messages and offers to shoppers while they are still in the store.

- **In a decision management framework, real-time actions are well-structured.** What constitutes a correct decision is predetermined to evoke the desired action by a person or a dependent application. Decisions are documented in programmable code, in a centralized application that ensures the decision is traceable all the way back to what operational data inputs were used. IDC has said that these coded, prescribed actions from decision management have power, and based on their research can have a “profound impact in creating competitive advantage”, as they “...shift from the "art of the decision" to the "science of the decision".

Turning on the Big Data Fire Hose

There has been much debate around ‘how big is big data?’ One truism is that it will continue to grow, according to IDC, at an annual rate of 40 percent over the coming decade. Streams of big data continually flow at rates of hundreds of thousands, to millions of events per second. These data streams originate from both machines and people, created by software powering this digital age. And given streams come from different origins it’s perhaps not surprising that

21 D. Vesset et al., IDC Identifies Decision Management as a Key Enabler of the Intelligent Economy, Jan 2011
http://www.idc.com/about/viewpressrelease.jsp?containerId=prUS22652411

they can contain different types of data, both unstructured and structured. Video feeds, authored social media chatter along with sensors and other transmitting device data, are all being continuously generated.

Individual readings, messages and alike also referred to as individual data objects (or events), tend to occur in succession – one reading from a mechanical sensor is transmitted, than another albeit multiple events can happen at the same time like with twitter feeds (at least when considering multiple tweeters). Regardless, these events are commonly bounded by time – a consistent scale upon which all types of events can typically be measured. Continuous, in their succession, marked by time and representing objects, encoding data or recording elements – these big data elements are known as event streams.

For many applications, there is a need to analyze data before it is out of date, before the opportunity is lost or the damage is done. For example, stopping a fraudulent transaction before the money is gone, being the first to grasp a unique trading opportunity, addressing an issue before it becomes a problem, and the list goes on. Given the volume and speed of event streams it quickly becomes impractical to store the data for analysis. Frankly, with millions of events per second, event streams would soon flood data lake banks.

To make sense of this data, and answer questions like: ‘Is this a good trading opportunity?’ or ‘Is this an illegitimate access to the electrical system?’ we need to change the traditional analysis paradigm. Instead of storing the data, we store the analysis – scrutinizing the data stream as it flows on by. And when event streams are processed continuously, decisions about the data importance can be made in real-time. Decisions such as:

- Is this clean data? If not, does it need to be cleansed? If it’s dirty, then clean it.
- Does this data fit a pattern that I care about? If so, take the necessary action (notify me, update the dashboard, alert the dependent applications, etc.). If not, do I want to throw it away?
- Should this data be stored for more in-depth (out-of-stream) analysis? If so, what new scenarios do I need to model? And which of these examine new patterns of interest that should be encoded back into the data streams?

In other words, with event stream processing, conditions can be assessed, cleansing and calculations can be done, and actions can be prescribed in real-time when the data is analyzed in-memory and in-stream.

---

Case Study: Continuous Operational Surveillance

- **Background:** Avoiding machine failure of high value assets, like the submersible oil pumps operating at the bottom of the ocean floor, requires ongoing assessment of conditions. Based on continuous readings from electrical and mechanical sensors (such as vibration, temperature and pressure), ongoing assessment of patterns that indicate potential breakdowns before they happen avoids non-production time, estimated at about two million dollars a day.

- **An event stream processing engine was embedded into the event stream containing encoded routines to cleanse data (when temperature readings went awry based on insignificant reasons, for example), calculate key metrics, and comparing the event data to previous conditions and thresholds. In-depth analysis of this live data provided the real-time situational awareness needed to minimize pump downtime. When trends indicated new conditional status emerging, the streaming engine could automatically deliver the data to the (out-of-stream) analysis environment which retained the record of situational history. Predictive analytics could then be used to judge if equipment was about to fail, determine the remaining lifetime of the pump or it’s sensor, and even define new, emerging patterns of interest to encode back into the event stream processing engine.

- **The scenario status, at any given point in time (or even over short periods of time), captured in event streams (as objects emanating from sensors, mobile devices and media streams, for example), provides an information context that enriches decisions and actions. Just as with enterprise decision management, these decision scenarios are defined in software code. Therefore, the policies and decision points (aka business rules) as well as the predictive models that describe the likelihood of future events are consistently executed when presented with the same input. They are traceable, exposed within the code itself, have a well-defined lineage, and can be centrally managed from a single interface. But unlike managed enterprise decisions, event stream processing reaches into places imperceptible to humans, providing a new set of eyes upon which prescriptive decisions can be made.

Conclusion

The growth and innovation issues that organizations currently encounter and at times, suffer from require treatment. In virtually every industry, understanding the symptoms of inefficiency, unresolved problems and lack of innovation is needed to be competitive in the era of big data. Many firms have already begun to fundamentally change how data is managed, taking advantage of new computing infrastructures. Big data analytics provides the necessary
situational assessment to understand what is occurring, what needs to be monitored, and what needs to change for future success.

Informed by facts and driven by data, controlled decisions assure the same actions are taken under the same scenario conditions. The variety of big data enriches the contextual understanding of scenarios, improving the decisions being made. Automation enables the tireless consistency of computers, which repeatedly take the appropriate action in a controlled, managed environment. When conditions change or are predicted to change, defined governance strategies and policies specify the corresponding action - and these too can be encoded into real-time, low-latency responses.

Regardless of scale, type, and speed of data, prescriptive analytics is not only possible, it may be the very thing needed to put analytics into action to solve organizational illness.

About the Author

Fiona McNeill is the Global Technology Product Marketing Manager at SAS. She has been described as a pioneer and one of the most influential people by CRMPower and is known as a speaker, author and innovator in the field of analytics. She has worked alongside some of the largest global organizations, helping them derive tangible benefit from the strategic application of analytics to real-world business scenarios. Over the course of her 16-year SAS career, she has worked in product management, product strategy, consulting, and sales. Currently, she is focused on SAS Text Analytics, SAS Event Stream Processing and SAS Decision Management. Prior to SAS, she was a member of IBM Global Services.

McNeill has recently published *Heuristics in Analytics* with co-author Carlos Andre Pinheiro, and has previously published both in academic and business journals. She received her M.A. in Quantitative Behavioral Geography from McMaster University, examining inter-temporal time dependence in consumer purchasing behavior and graduated cum laude with a B.Sc. in Bio-Physical Systems, University of Toronto.

Additional Information

To learn more about this topic, please visit [www.sas.com/mandate](http://www.sas.com/mandate) on [sas.com](http://sas.com)