OCTOBER 2014

TDWI E-Book

Predictive Analytics: Revolutionizing Business Decision Making

Sponsored by:

§sas®

tdwi.org

tdwi

PREDICTIVE ANALYTICS 101

Exploring predictive analytics for the first time? Malene Haxholdt, global marketing manager for business analytics at SAS, explains what you need to know to get started.

TDWI: From your perspective, what is the value of predictive analytics?

Malene Haxholdt: Companies are looking to get value from predictive analytics in many business areas. The value comes when you can take data, apply analytics, and act on the results. The value ultimately means growing revenue, lowering costs, or establishing governance and compliance. Growing revenue is often associated with customer analytics and being better at retention modeling and cross-sell and up-sell activities. Lowering costs often comes from analytics used to improve processes as well as from using internal capacity in any form. Especially in the financial industry, the driver for predictive analytics can be part of a compliance and governance strategy.

Is predictive analytics used more often in certain industries?

Predictive analytics is relevant and useful across all industries. Some industries are more mature in their use and implementation of predictive analytics. The most common applications of predictive analytics come from a need to better manage fraud, risk, equipment failure, or customer interactions. What are common across industries are the accelerating growth of data and the desire to turn it into valuable information. Predictive analytics is a component of that journey.

What are the top things to consider if you want to implement predictive analytics?

It is key to remember that predictive analytics is only valuable if you can turn the results of analytical models into actions. The business process, the people, and the technology need to be aligned to successfully deploy predictive analytics. Think of predictive analytics as part of a life cycle that consists of (1) managing all of your data, (2) exploring all of your data, (3) building your models with the best techniques, and (4) deploying and monitoring your models.

What human skill sets are needed to be successful with predictive analytics?

The important part is that the organization believes and understands that predictive analytics is driving better decisions. The software tools that allow you to start building predictive models are very approachable and do not require you to be a technical expert. We see a tremendous uptick in interest in interactive predictive modeling using visual analytics and visual statistics. Without any coding required, you can start exploring your data and get value. More mature companies are hiring more data scientists to further grow and explore the possibilities predictive analytics brings. Having human skills that can align the technology and business understanding is key to success.

What new predictive analytics techniques are companies exploring?
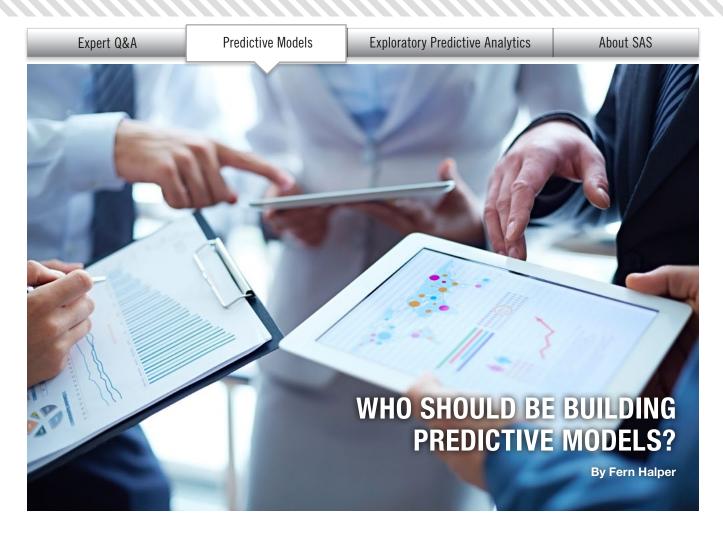
With the reality of big data, new techniques are being explored by companies to leverage the value hidden in new types of data. Being able to explore *all* of your data quickly and in an interactive manner is driving the need for data visualization techniques and interactive predictive modeling on very big amounts on data—fast. Also, we see a growing interest in machine-learning techniques such as Random Forest and techniques to handle unstructured data. Take, for example, SAS Contextual Analysis. This next-generation text modeling software combines the ease of machine learning with subject-matter expertise, enabling powerful text models to easily be defined from unstructured data.

In the world of predictive analytics, what is IT's role?

To get sustainable value from predictive analytics, IT and business users are both key in the process. A modern CIO and IT department work closely with the business to enable predictive analytics throughout the organization by providing data access and approachable analytics tools to the right users. Think of it as self-service analytics. The IT department is, in some cases, the true driver of innovation because it can enable the use of all data and help get even more predictive power in the models used by the business. IT's role is critical in selecting an architecture that will meet future demands around diverse analytical data preparation tasks, reducing model building latency and quickly deploying models into operational systems.

Should I care about predictive analytics if I work for a small company?

The short answer is yes! All companies, no matter the size, will benefit from making better decisions in a world of uncertainty. We see small companies that apply predictive analytics as a key strategy to growth. In fact, companies that embed predictive analytics as a cornerstone in their business tend to be more successful in the long run.

WHO SHOULD BE BUILDING PREDICTIVE MODELS?

By Fern Halper

With statistical and data mining skills in short supply, are the right people building your models? These three safeguards can help you avoid model-building problems.

There has been quite a bit of talk recently about the democratization and consumability of analytics. *Democratization* refers to extending the deployment of BI and analytics tools to more users in an organization. The idea is to let all people, regardless of technical prowess, have access to data to analyze and enable more informed decisions. Democratization originally focused on more descriptive analytics and reports and dashboards. Now this movement includes visualization as well as more advanced techniques. *Consumability*, which is related to democratization, means either that BI or analytics can be used easily by a lot of people, or that the results of BI or analytics can be consumed by the masses. I'm referring to the latter here.

Concurrent with these market trends is the movement to empower business analysts to use more advanced forms of analytics, such as predictive analytics. Is this a good idea? The market seems to think so. In a 2014 TDWI Best Practices Report on predictive analytics, 86 percent of respondents currently using predictive analytics expected that business analysts would be the primary builders of predictive models in

the near future. Seventy-nine percent said that statisticians or other quantitative types would be the primary builders. Additionally, when asked what the top three skills for model building were, respondents chose knowledge of the business, knowledge of the data, and critical thinking. Statistical knowledge and knowledge of the tool ranked low.

What's going on here? Is this a push by the vendors to put their tools into more hands, or a movement by the business to do so? It is probably a bit of both.

On the one hand, vendors have made their software easier to use. They are including drag-and-drop interfaces rather than requiring the model builder to use a scripting language. Some software tools suggest models based on input data and the specification of targets of interest, such as *buy* or *don't buy*. This has made the software easier to use. On the other hand, statistical and data mining skills are in short supply. Some organizations feel that their highly skilled staff can work together with business analysts to build models, especially if they understand the data and can deal with it.

The question still remains: Is this a good idea, or will we be reading a few years from now that someone who was not qualified built a predictive model that was put into production and cost a company millions of dollars?

Three safeguards can help you avoid that situation.

Training: I don't believe that just any business analyst can build a predictive model, even a critical thinker with knowledge of the business and the data. Training is necessary for any kind of predictive analysis for two reasons.

First, it is important to understand what you are doing, especially if you have to defend your analysis. Second, it makes sense to get training on a tool you are using so you can use it correctly. There are pitfalls to predictive modeling: overfitting, giving irrelevant data too much credence, or thinking that correlation means causation. It might involve not understanding how models get stale or how to treat certain kinds of data or build the right attributes for the task. If you're building predictive models and don't understand the last two sentences, then my point is made. I'm not saying you need a Ph.D. in statistics, but data mining boot camps as well as online courses or vendor-provided training can help increase your knowledge.

Collaboration: A company might let a business analyst build a predictive model, but in order for it to get into production to make decisions, an enterprise needs a control process. This process might involve the business analyst getting approval from the data scientist for any model that the business analyst builds. It might include business analysts and data scientists working together on the models. Vendors are helping by providing collaboration and sharing features in their toolkits.

Risk analysis: An enterprise needs to decide for itself what kinds of models it might let its business analysts create. The data scientist or statistician might be the only person allowed to build models that carry a high-risk/high-dollar tag. A business analyst might be able to create less risky models. It is up to the organization to decide.

Conclusion

Just because something is easier to do doesn't mean everyone should be doing it. Who builds models should be taken seriously. However, if companies take precautions, they may be able to take the first steps toward opening up predictive analytics to more builders.

Fern Halper, Ph.D., is well known in the analytics community, having published hundreds of articles, research reports, speeches, Webinars, and more on data mining and information technology over the past 20 years. Halper is also co-author of several "Dummies" books on cloud computing, hybrid cloud, and big data. She is the director of TDWI Research for advanced analytics, focusing on predictive analytics, social media analysis, text analytics, cloud computing, and "big data" analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her at fhalper@tdwi.org, on Twitter @fhalper, and on LinkedIn at linkedin.com/in/fbhalper.

EXPLORATORY PREDICTIVE ANALYTICS: BOOST MODEL DEVELOPMENT AND PRODUCTIVITY USING A VISUAL PARADIGM

Technology has finally started to catch up with the needs of business analysts, marketing data scientists, and other domain experts. The future of exploratory predictive analytics consists of an interactive, data-visualization-powered user experience with powerful behind-the-scenes in-memory analytics. Tapan Patel, product marketing manager at SAS, provides some perspective.

Can predictive analytics ever truly be made pervasive? It sure can!

In the end, it comes down to what we mean by "pervasive." Can information workers in an organization *consume* and *benefit* from predictive analytics (PA)? Absolutely. In the final analysis, this is precisely what's meant by making predictive analytics truly "pervasive."

Face it: The average employee probably doesn't have much interest in building, testing, and training a predictive model. Zoe in online sales just wants to cross-sell or up-sell products to customers and meet her targets. On the other hand, Zoe is uniquely positioned to benefit from predictive insights that permit her to understand a given customer interaction or behavior across different channels, recommend products that can generate the best revenue opportunities, or make time-critical adjustments to retain high-value customers. Making PA pervasive means pushing it out to information consumers such as Zoe. It means *operationalizing* predictive analytics (i.e., embedding insights into business processes, where they can provide the greatest benefit).

Making PA pervasive also means solving the challenge on the "producing" end—that there's a small pool of professionals

who are savvy enough in statistics, data mining, and business domain expertise to design and build predictive models. There is a need to expand the "producers" pool and let analytically "new" users, business analysts, and functional domain experts create and refine models. Technology has finally caught up with the needs of these types of end users.

Predictive analytics that combines user self service and an interactive, data-exploration-powered user experience with behind-the-scenes in-memory processing will change the status quo. This new take doesn't enlist Zoe and others like her as PA producers; it *does*, however, recast more staff members as producers of predictive analytics insights.

"We're seeing a class of users, including marketing data scientists, risk analysts, and product/brand managers—people responsible for segmentation, campaigns, credit risk, customer propensity, and next-best action—involved in building predictive models and collaborating with IT and data management staff to put it into production," says Tapan Patel, product marketing manager at SAS.

How is this possible? As Exhibit A, Patel points to SAS Visual Statistics, a new generation of easy-to-use, interactive data discovery environment designed *expressly* for data exploration and model development. Both of these go hand-in-hand to find out why certain events happen, how strong the relationships are, and whether they could be exploited to predict future outcomes.

"How can I interact with variables to quickly add the relevant ones, remove outliers, and instantly visualize the results as part of a model refinement process?" he asks. Users can ask for more real-time interactions from the data, identify the root cause at observation level, and reduce the feature selection and model-building latencies.

Patel's Exhibit B is under the covers. The interactivity, fast response times, and multi-user concurrency is affordable because of the distributed in-memory analytics platform. "Once the required data is loaded in memory, users can make multiple passes through the data for analytical computations or actions," he says. Because the data resides in memory, users can quickly build numerous models by group or segment (e.g., location, store, owner, device, age, income) on the fly to exploit specific relationships to the group without having to sort or index the data each time.

Eggheads Matter Now More Than Ever

The advent of this new class of interactive predictive analytics tools doesn't mean that data scientists, statisticians, and other super-savvy users will suddenly cease to be valuable. Far from it.

In most cases, Patel argues, experts will be freed up to focus on high-value-added areas such as:

- Continuing to perform exploratory or discovery analysis to find unknown opportunities or risks (which can lead to high-value discoveries for use in downstream analytical projects)

- Developing sophisticated predictive models using the latest machine learning techniques and ensembles

- Overseeing the vetting, training, and improvement of the models or insights developed by business analysts and less-savvy users

- Devising creative solutions to problems, be it by bringing in new sources of data (e.g., from outside or subscription sources) and variables, developing or selecting new algorithms and functions, applying a rigorous approach to problem solving, or all of the above

Analytical Data Preparation Is a Critical Step in Your Predictive Analytics Journey

"Irrespective of the analytical skill set or prowess, any model development task will require a lot of time spent in data preparation activities. Analytical data preparation is unique in the sense that each modeler or statistician or business analyst might have a different business problem to solve, and it requires a lot of iterative and collaborative effort and in general is a very time-consuming step. The creativity element is also required—the ability to look at data from different angles and manipulate it quickly, handle missing values or variable reduction techniques, drop variables that are not useful, and bring in unique, new variables to improve outcomes or results from the modeling exercise. This can't be taught overnight. Besides a comprehensive set of tooling, a

thorough understanding of your problem and data elements needed to solve it is equally important."

You Have to Deploy Your Best Models and Make Sure They Do Not Become Stale

Making the analytical life cycle steps efficient, governed, and automated is also crucial, both with respect to the background data preparation work that feeds analytical processes (as discussed earlier) *and* to the deployment of analytical models against operational data. Vendors are now addressing both of these important steps in earnest.

Automating the model deployment step is equally important to generate timely insights. The step should be augmented by combining analytical models and business rules for operational decisions and continuous monitoring of models for relevancy and accuracy.
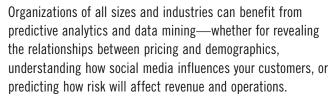
"With the onslaught of big data and need to make many high-volume, frontline decisions, the question of how to quickly apply the best analytical models to new data will become even more important," Patel says. These operational types of decisions (e.g., customer retention, risk underwriting, fraud) require a combination of business rules, analytics, and model management to prescribe consistent and accurate insights. Still another critical area of focus has to be ongoing monitoring of the health of your models.

Predictive models essentially embody a representation of a business "world" at a certain point in time; as a result, they'll tend to grow "stale" with rapid changes (e.g., new pricing, new competitors, new customer preferences). Fixing, retraining, scrapping, or building new models will continue to be a big part of the data scientist's job, Patel argues. "This is where the concept of model performance monitoring becomes critical."

SAS is designed to support the entire analytics life cycle process—from data preparation and exploration to model development and deployment, Patel notes. Technology isn't always the only silver bullet, however. Cooperation and collaboration between and among analytic professionals, business stakeholders, IT staff, and others is of paramount importance. It's in everyone's best interests that the right data is made available at the right time so that models can be thoroughly developed, validated, and deployed; it's no less important that high-value predictive models, once validated, are embedded in operational applications. Software isn't everything; people and processes are equally important to success.

sas.com

Organizations of all sizes and industries can benefit from predictive analytics and data mining—whether for revealing the relationships between pricing and demographics, understanding how social media influences your customers, or predicting how risk will affect revenue and operations.

With an unwavering focus on data analytics since 1976, SAS offers a broad set of predictive analytics and data mining solutions with a full range of on-site and cloud-based deployment options.

SAS® Data Mining allows you to easily prepare diverse data sets for analytics and quickly engage in data discovery. It includes a broad set of descriptive, predictive, and machine learning algorithms and a mechanism to automatically deploy and manage a growing number of models to ensure relevance and usefulness.

SAS continues to lead the predictive analytics market with the widest set of product capabilities and is frequently selected based on quality, scalability, model efficiency against wide data sets, and large end-user community. SAS is also meeting challenging customer requirements around Hadoop, streaming data, real-time analytics, and self-service analytics.

Learn more on some of the hot topics in analytics at sas.com/en_us/insights/analytics.html.

tdwi.org

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence, data warehousing, and analytics industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, on-site courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive website, tdwi.org.