

Analytics

DRIVING BETTER BUSINESS DECISIONS

NOVEMBER/DECEMBER 2015

BROUGHT TO YOU BY:

informatics

REALITY OF REAL-TIME ANALYTICS

Four important areas where real-time analytics can provide real value

ALSO INSIDE:

- Big data in marketing analytics
- Mother Nature meets Internet of Things
- Simulated worlds: Simulation software
- Data reduction for everyone



Executive Edge

Mike Neal, CEO of DecisionNext, on which makes better business decisions: humans or machines?

Humans vs. machines

In his executive edge column in this issue of *Analytics*, Mike Neal, CEO of DecisionNext, poses an interesting question: “Have we grown too dependent on technology, specifically automation that makes important decisions for us?”

The decade-old analytics wave – built on the premise that data-driven decision-making is superior to good, old-fashioned gut instinct – has swept over the business world. Clearly, there is no going back as industry after industry has reaped the benefits of automation and analytics. But is there still a role for humans in the decision-making process?

Despite their biases and blind spots, humans are pretty good when it comes to nuance, the art of the deal. Writes Neal: “Computers are great at finding connections between data points, while humans are great at assigning meaning to those connections. Human experts in any given industry, therefore, are well suited for applying predictive analytics to real-world decisions.” For more on the story, see [“Humans or machines.”](#)

Speaking of humans and human resources, Chandrakant Maheshwari, a subject matter expert in the risk and regulatory practice at Genpact LLC, praises the power of human curiosity in his Forum column titled [“Driving curiosity as a culture in an analytics organization.”](#)

Writes Maheshwari: “Curiosity and continuous learning have become critical qualities for the analytics professional, but the biggest challenge is knowing what to learn and where to explore.”

For our cover story, Larry Skowronek, senior vice president of product management for Nexidia, takes a look at what he describes as “one of the most interesting trends in the analytics space today”: real-time analytics. In his article [“The reality of real time,”](#) Skowronek presents four key areas – at-risk compliance, customer retention, increased sales and post-call analytics – where real-time analysis and interaction with customers results in the “best possible” business outcomes.

As I’m writing this a week before Halloween, I’m also preparing to travel to Philadelphia for the INFORMS Annual Meeting on Nov. 1-4. I’m looking forward to attending many of the technical sessions, keynote presentations and social events, along with seeing some longtime friends like Vijay Mehrotra, entrepreneur/professor and author of the “Analyze This!” column. In his latest [column](#) Vijay recalls his first trip to an INFORMS conference 25 years ago, also in Philadelphia, and how the conference and a brief chat changed his career path. **|**

– PETER HORNER, EDITOR
peter.horner@mail.informs.org



Experience Optimization Like Never Before
More Scenarios, Smarter Evaluations, Better Decisions



WHY CHOOSE AIMMS PRO?

Discover how we can meet the unique needs of your business.



User Friendly

Interactive dashboards in a web browser provide analytics and decision support organization-wide.



Powerful

Robust modeling and optimization helps you see options, compare results and chart your course.



Configurable

Create custom apps – from Network Optimization to Portfolio Management, Pricing and more.



Easy to Update

Make changes centrally, distribute updates instantly across your organization - no more spreadsheets.



Seamless

Effortlessly connect with your existing ERP, spreadsheets and other systems to maximize data.



Access Anywhere

Solve problems on the go and respond in real time from any device.

Get your FREE AIMMS 30 Day Trial

Email us at info@aimms.com or call 425-458-4024



Award winning software company AIMMS works with half of the Global Fortune Top 20 companies to optimize their businesses, making them even more successful.



Analytics

DRIVING BETTER BUSINESS DECISIONS

NOVEMBER/DECEMBER 2015

Brought to you by




40

FEATURES

34 THE REALITY OF REAL TIME

Marketing and customer interaction: Four important areas where real-time analytics can provide real value.

By Larry Skowronek

40 BIG DATA IN MARKETING ANALYTICS

Having big data doesn't automatically lead to better marketing. It's the insights derived and decisions made.

By Mousumi Ghosh

44 SIMULATED WORLDS

Biennial simulation software survey: Driven by questions, fueled by thought and realized by simulation.

By James J. Swain

54 MOTHER NATURE MEETS IOT

What does the oil and gas industry have in common with agriculture? Much more than you think.

By Atanu Basu and Gabe Santos

58 DATA REDUCTION FOR EVERYONE

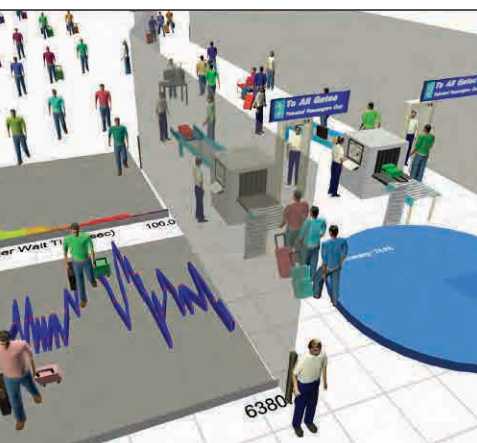
Distillation of multitudinous data into meaningful parts plays key role in understanding world around us.

By Will Towler

66 'ENSEMBLE OF ENSEMBLE'

Increased focus on accuracy in predictive modeling results in newer and more innovative techniques.

By Anindya Sengupta



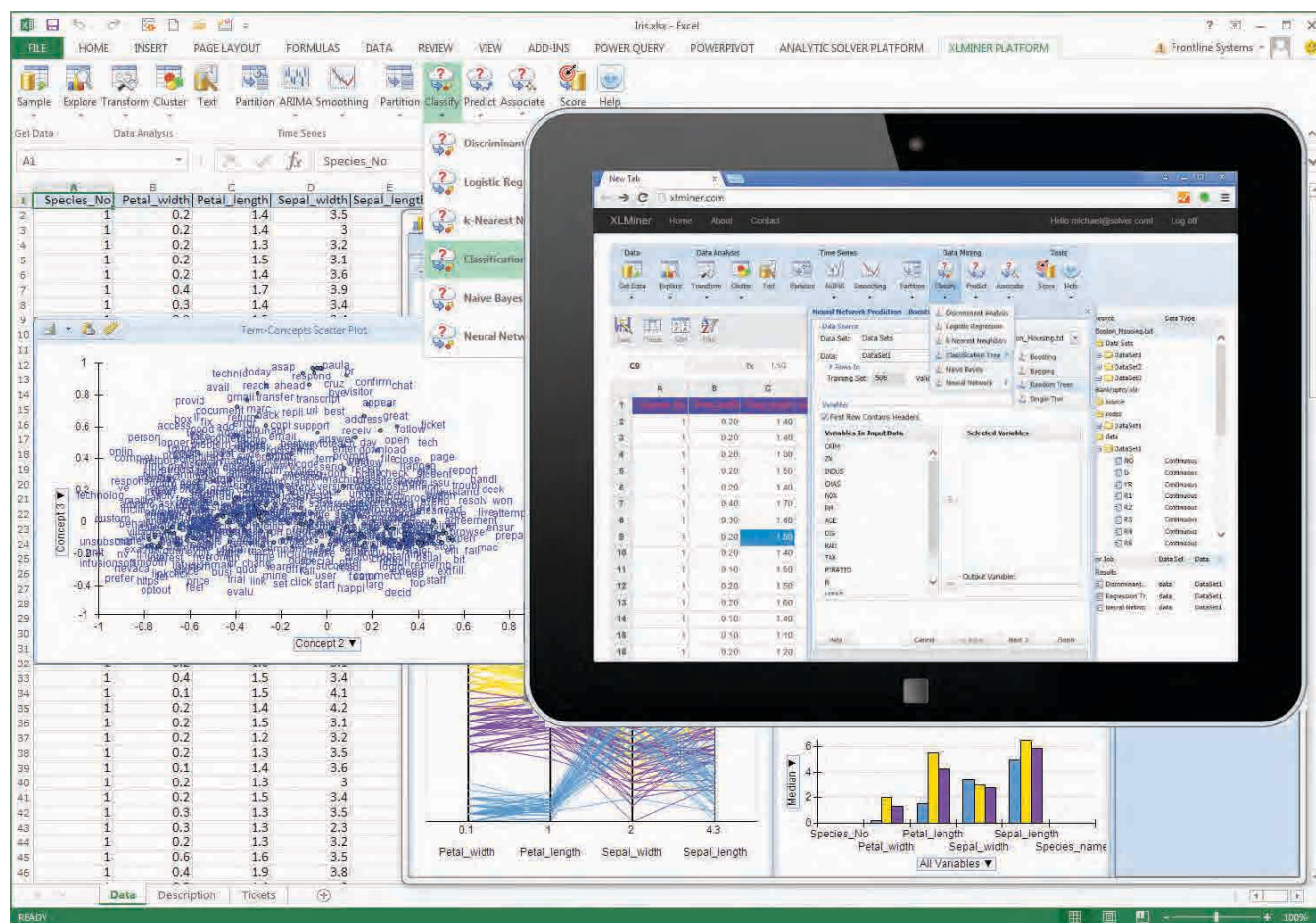
44



54

XLMiner®: Data Mining Everywhere

Predictive Analytics in Excel, Your Browser, Your Own App



XLMiner® in Excel – part of **Analytic Solver® Platform** – is the most popular desktop tool for business analysts who want to apply data mining and predictive analytics. And soon it will be available on the **Web**, and in **SDK** (Software Development Kit) form for your own apps.

Forecasting, Data Mining, Text Mining in Excel.

XLMiner does it all: Text processing, latent semantic analysis, **feature selection**, principal components and clustering; exponential smoothing and ARIMA for **forecasting**; multiple regression, k-nearest neighbors, and ensembles of regression trees and neural networks for **prediction**; discriminant analysis, logistic regression, naïve Bayes, k-nearest neighbors, and ensembles of classification trees and neural nets for **classification**; and association rules for affinity analysis.

XLMiner.com: Data Mining in Your Web Browser.

Use a PC, Mac, or tablet and a browser to access all the forecasting, data mining, and text mining power of XLMiner **in the cloud**. Upload files or access datasets already online. Use the same Ribbon and dialogs you

have in Excel, and generate the same reports, displayed in your browser or downloaded for local use.

XLMiner SDK: Predictive Analytics in Your App.

Access all of XLMiner's parallelized forecasting, data mining, and text mining power in your own application written in **C++**, **C#**, **Java** or **Python**. Use a powerful object API to create and manipulate DataFrames, and combine data wrangling, training a model, and scoring new data in a single operation "pipeline".

Find Out More, Start Your Free Trial Now.

Visit www.solver.com to learn more, register and download Analytic Solver Platform or XLMiner SDK. And visit www.xlminer.com to learn more and register for a free trial subscription – or email or call us today.



24



70

DEPARTMENTS

- 2 Inside Story
- 8 Executive Edge
- 14 Analyze This!
- 20 Healthcare Analytics
- 24 INFORMS Initiatives
- 28 Forum
- 32 Newsmakers
- 70 Five-Minute Analyst
- 74 Thinking Analytically

Analytics (ISSN 1938-1697) is published six times a year by the Institute for Operations Research and the Management Sciences (INFORMS), the largest membership society in the world dedicated to the analytics profession. For a free subscription, register at <http://analytics.informs.org>. Address other correspondence to the editor, Peter Horner, peter.horner@mail.informs.org. The opinions expressed in *Analytics* are those of the authors, and do not necessarily reflect the opinions of INFORMS, its officers, Lionheart Publishing Inc. or the editorial staff of *Analytics*. *Analytics* copyright ©2015 by the Institute for Operations Research and the Management Sciences. All rights reserved.

Analytics

DRIVING BETTER BUSINESS DECISIONS

informs

REGISTER FOR A FREE SUBSCRIPTION:

<http://analytics.informs.org>

INFORMS BOARD OF DIRECTORS

President	L. Robin Keller, University of California, Irvine
President-Elect	Edward H. Kaplan, Yale University
Past President	Stephen M. Robinson, University of Wisconsin-Madison
Secretary	Brian Denton, University of Michigan
Treasurer	Sheldon N. Jacobson, University of Illinois
Vice President-Meetings	Ronald G. Askin, Arizona State University
Vice President-Publications	Jonathan F. Bard, University of Texas at Austin
Vice President-Sections and Societies	Esma Gel, Arizona State University
Vice President-Information Technology	Marco Lübbecke, RWTH Aachen University
Vice President-Practice Activities	Jonathan Owen, CAP, General Motors
Vice President-International Activities	Grace Lin, Institute for Information Industry
Vice President-Membership and Professional Recognition	Ozlem Ergun, Georgia Tech
Vice President-Education	Jill Hardin Wilson, Northwestern University
Vice President-Marketing, Communications and Outreach	E. Andrew "Andy" Boyd, University of Houston
Vice President-Chapters/Fora	David Hunt, Oliver Wyman

INFORMS OFFICES

www.informs.org • Tel: 1-800-4INFORMS

Executive Director	Melissa Moore
Meetings Director	Laura Payne
Director, Public Relations & Marketing	Jeffery M. Cohen
Headquarters	INFORMS (Maryland) 5521 Research Park Drive, Suite 200 Catonsville, MD 21228 Tel.: 443.757.3500 E-mail: informs@informs.org

ANALYTICS EDITORIAL AND ADVERTISING

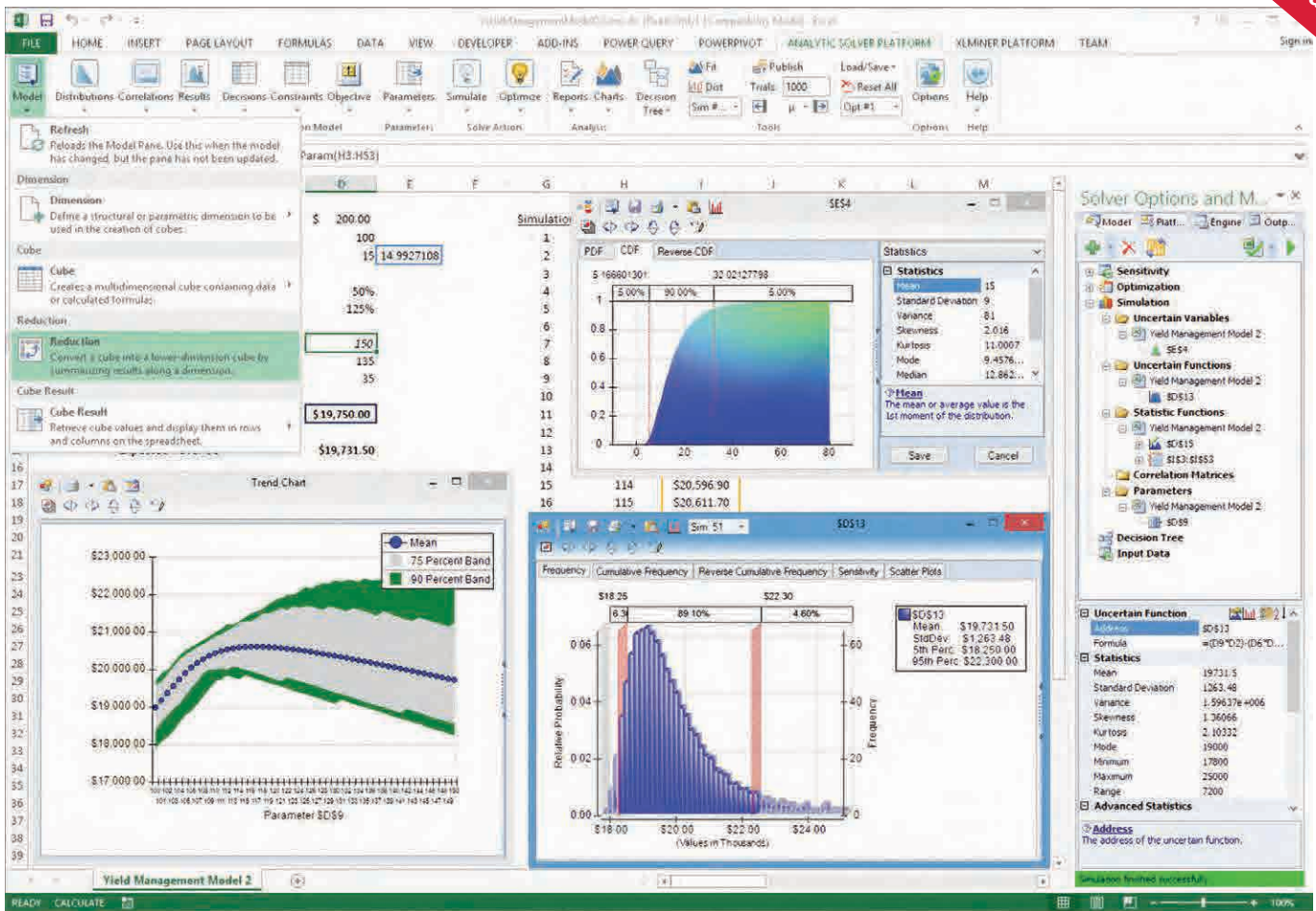
Lionheart Publishing Inc., 506 Roswell Street, Suite 220, Marietta, GA 30060 USA
Tel.: 770.431.0867 • Fax: 770.432.6969

President & Advertising Sales	John Llewellyn john.llewellyn@mail.informs.org Tel.: 770.431.0867, ext. 209
Editor	Peter R. Horner peter.horner@mail.informs.org Tel.: 770.587.3172
Assistant Editor	Donna Brooks donna.brooks@mail.informs.org
Art Director	Jim McDonald jim.mcdonald@mail.informs.org Tel.: 770.431.0867, ext. 223
Advertising Sales	Sharon Baker sharon.baker@mail.informs.org Tel.: 813.852.9942

ANALYTIC SOLVER® PLATFORM

From Solver to Full-Power Business Analytics in Excel

Supports Tableau,
Power BI and
Apache Spark Big Data



Solve Models in Desktop Excel or Excel Online.

From the developers of the Excel Solver, Analytic Solver Platform makes the world's **best optimization software** accessible in Excel. Solve your existing models faster, scale up to large size, and solve new kinds of problems. Easily **publish** models from Excel to share on the Web.

Conventional and Stochastic Optimization.

Fast linear, quadratic and mixed-integer programming is just the starting point in Analytic Solver Platform. Conic, nonlinear, non-smooth and global optimization are just the next step. Easily **incorporate uncertainty** and solve with simulation optimization, stochastic programming, and robust optimization – all at your fingertips.

Fast Monte Carlo Simulation and Decision Trees.

Analytic Solver Platform is also a full-power tool for Monte Carlo simulation and decision analysis, with 50 distributions, 40 statistics, **Six Sigma** metrics and risk measures, and a wide array of charts and graphs.

Plus Forecasting, Data Mining, Text Mining.

Analytic Solver Platform samples data from Excel, PowerPivot, and SQL databases for forecasting, data mining and text mining, from time series methods to classification and regression trees and neural networks. And you can use visual data exploration, cluster analysis and mining on your *Monte Carlo simulation* results.

Find Out More, Download Your Free Trial Now.

Analytic Solver Platform comes with Wizards, Help, User Guides, 90 examples, and unique **Active Support** that brings live assistance to you right inside Microsoft Excel.

Visit www.solver.com to learn more, register and download a free trial – or email or call us today.



FRONTLINE
solvers®

The Leader in Analytics for Spreadsheets and the Web
Tel 775 831 0300 • info@solver.com • www.solver.com

Humans or machines

Which makes better business decisions?
(Spoiler alert: both)

Computers are great at finding connections between data points, while humans are great at assigning meaning to those connections.



BY MIKE NEAL

Within the past year, a series of unrelated technical glitches brought the New York Stock Exchange (NYSE) to a grinding halt for more than three hours, delayed hundreds of United Airlines flights and took down the *Wall Street Journal's* home page.

Experts agree the amount of damage done was relatively minor, but the episodes reopened the issue of whether we have [grown too dependent on technology](#), specifically, automation that makes important decisions for us.

There's no doubt automation is a boon in many respects. From [next-day package delivery](#) to ATMs to safer industrial workplaces, automation definitely makes our lives easier and safer in a number of ways.

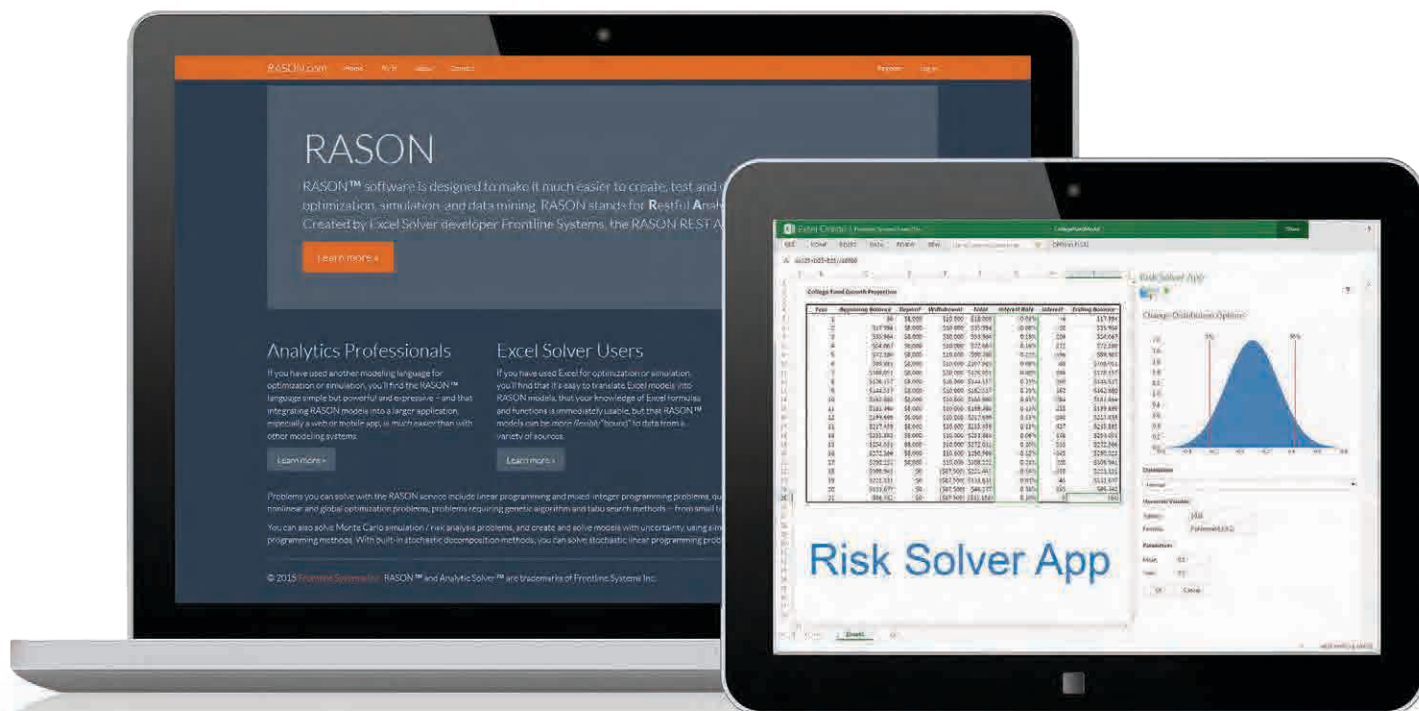
Some even claim that [automation can save local jobs](#) by helping factories remain cost-competitive with cheaper overseas labor. In many cases, computers are able to outperform humans in tasks such as [legal discovery](#), [making medical diagnosis](#) and [evaluating the potential of a job candidate](#).

On the other hand, increased automation creates new risks for society. United Airlines, for instance, pointed to an "automation issue" to explain its recent system-wide shut down.

ANALYTICS IN YOUR BROWSER

Optimization, Simulation, Statistics in Excel Online, Google Sheets – or Your Own Web or Mobile Application with RASON® Software

RASON® v2 with
Powerful Indexing,
QData Support



Advanced Analytics built on **Business Intelligence** is moving to the Web, and getting easier! Spreadsheet users can get results quickly in **Excel Online** or **Google Sheets**. Analysts and developers can build Web and mobile apps quickly using our new **RASON®** language.

Use our Analytics Apps in your Web Browser.

Solve linear, integer and nonlinear optimization models with Frontline's free **Solver App**, for desktop Excel Solver features in Excel Online and Google Sheets.

Run Monte Carlo simulation models with Frontline's **Risk Solver® App** – free for limited-size models, and free without size limits for desktop Risk Solver Pro, Risk Solver Platform, and Analytic Solver® Platform users.

Use our free **XLMiner® Analysis ToolPak App** for statistical analysis in Excel Online and Google Sheets, matching the Analysis ToolPak in desktop Excel – and our free **XLMiner® Data Visualization App** to explore your data with multivariate charts in Excel Online.

Build Your Own Apps with RASON® Software.

RASON – RESTful Analytic Solver® Object Notation – is a new modeling language for optimization and simulation that's embedded in **JSON** (JavaScript Object Notation).

With support for linear, nonlinear and stochastic optimization, array and vector-matrix operations, and dimensional tables linked to external databases, the **RASON** language gives you all the power you need.

Application Deployment Made Easy.

The magic begins when you deploy your model to the Web. Unlike other languages, **RASON** models are valid in JavaScript. Your browser or mobile apps can access our **cloud servers** via a simple **REST API**, to easily solve the largest models on the smallest devices. And your server-based applications can easily solve RASON models with our **Solver SDK®** object-oriented library.

Find Out More, Sign Up for a Free Trial Now.

Visit www.solver.com/apps to learn more, and visit rason.com to sign up for a free trial of RASON and our REST API. Or email or call us today.

 **FRONTLINE solvers®**
The Leader in Analytics for Spreadsheets and the Web
Tel 775 831 0300 • info@solver.com • www.solver.com

Should there be a limit to how much automation we accept into our lives, or should we hand more and more tasks to computers, willy-nilly?

On a larger scale, many experts say it was [faulty predictive modeling](#) that caused the financial crisis of 2008. Statistical models said it was unlikely that a critical mass of homeowners would default on their mortgages all at the same time, but the models were built with one bad assumption, so they did default – and an epic downturn of mortgage-backed securities caused a global financial crisis.

So what is the answer? Should there be a limit to how much automation we accept into our lives, or should we hand more and more tasks to computers, willy-nilly?

It's a false dichotomy.

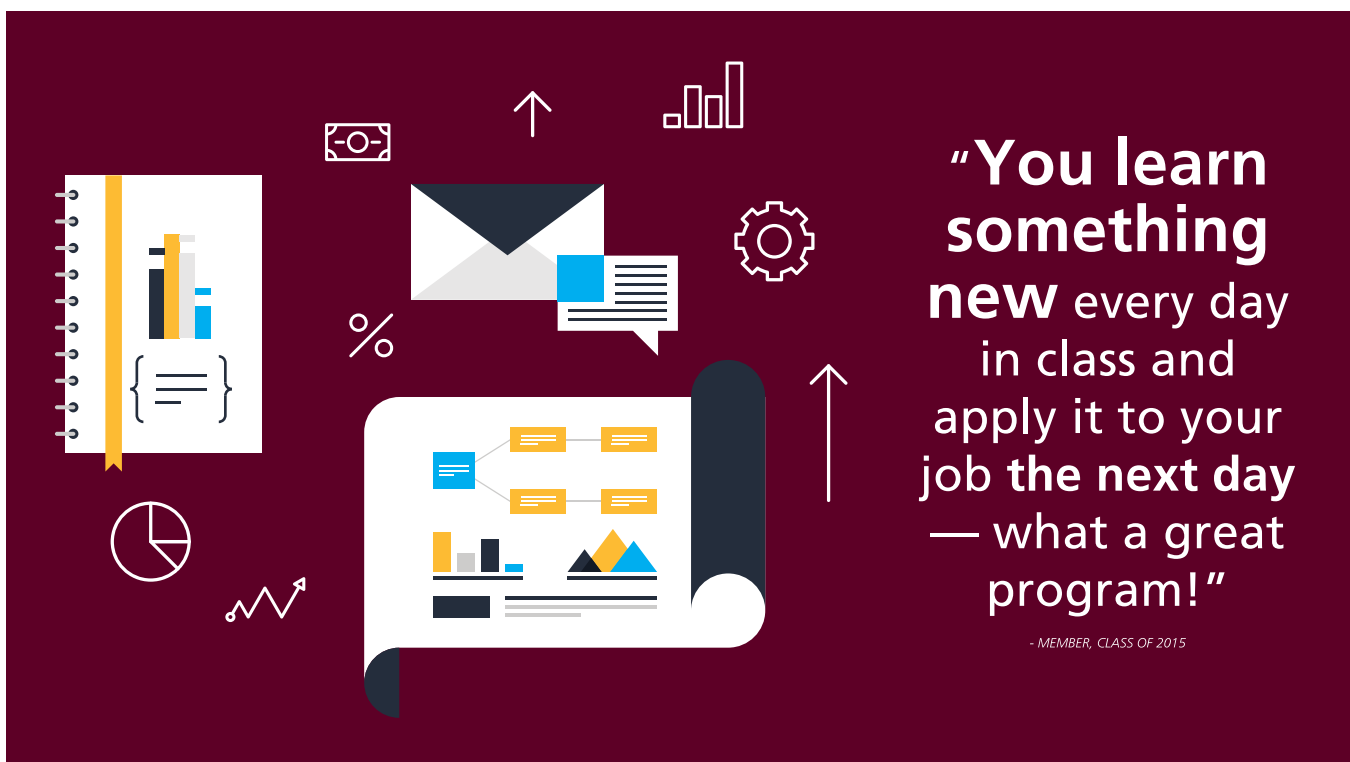
Not only have automated technologies and predictive modeling been around for [hundreds of years](#), but when it comes to some business applications such as predicting market behavior, it's actually a combined approach that reliably leads to the strongest results. Not only can computer models bring scale and speed, but they can help remove human biases from our decisions. Human beings bring insight, imagination and subtler forms of pattern recognition.

A well-known example happened in 2005 when two chess amateurs – [Steven Cramton and Zackary Stephen](#) – beat several grand masters as well as Hydra, the most advanced supercomputer at the time (more powerful than Deep Blue). They did it through a combination of their own knowledge and computer models, proving that humans working with technology can be smarter than either technology or humans alone.

So, blending these two decision-making mechanisms can actually lead to better results than either could produce independently. For example, automated models can actually measure and calibrate human analysts' predictions to find the most reliable experts and adjust their estimates to account for their relative accuracy and unique biases. This "model of models" can be more accurate than the humans or machines alone.

Another example would be to allow an experienced human analyst to override certain key inputs to a predictive model, so that a model can take into account breaking news, something difficult to capture in a fully automated model. So, in the same way that amateur chess players depend on a computer's insights, machines can do the same with human intelligence.

In this sense, automation and algorithms aren't about technology replacing



"You learn something new every day in class and apply it to your job the next day — what a great program!"

- MEMBER, CLASS OF 2015

**ANALYTICS**
TEXAS A&M UNIVERSITY

75% STATISTICS + 25% BUSINESS = 100% COMPETITIVE
analytics.stat.tamu.edu

Human decision-making is limited by complexity and human biases, while predictive models are only as good as the assumptions that go into them.

human beings or human intelligence; they're about technology being used to assist and inform human intelligence, especially with decision-making. In fact, used correctly, predictive analytics not only doesn't threaten human skills or destabilize markets, it may have the power to make markets more stable and predictable.

Consider commodities markets. They are notorious for price volatility and difficult to time. With the marketplace populated by more players than the actual producers or consumers of the physical commodity, it is a prescription for volatility. When you combine that volatility with the complexity of deciding among various processing options, it is easy to understand how predictive models might play an important role. The most profitable decisions can easily be missed when "gut instinct" is the decision-making approach.

On top of that, highly trained commodities experts often end up spending too much of their time chasing down data. Predictive modeling holds the promise of giving these experts more time to do what they do best: analysis.

Consider that in the 1970s and 1980s, American Airlines used predictive technologies to increase its market share even as nine other major airlines and hundreds of smaller carriers went out of business. They did this by introducing sophisticated new data-driven pricing strategies. This "revenue management" science used models to maximize capacity on each flight while also maximizing revenue per seat. This was pricing optimization for a highly perishable

commodity, since when the plane pulls away from the gate the value of an unsold seat falls to zero. If predictive analytics can do that for an airline, what can it do for commodities markets? Most of these industries aren't using predictive analytics yet. Could predictive analytics help commodities and maybe even economies become more stable?

One thing's for certain: Computers are great at finding connections between data points, while humans are great at assigning meaning to those connections.

Human experts in any given industry, therefore, are well suited for applying predictive analytics to real-world decisions. Remember the amateur chess players? They beat the super-computer and grandmasters alike by taking information from a computer then filtering it through their own human decision-making.

Nate Silver has argued that data-driven decisions aren't actually about making predictions at all – they're about probabilities. As human beings, we crave certainty; we over-weight a prediction, and we tend to under-weight the error bars around it. But when we depend on computers to give us certainty, bad things can happen, particularly when we try to apply automation and prediction on a grand scale.

Our human decision-making is limited by complexity and human biases, while predictive models are only as good as the assumptions that go into them. The key is to combine the best of both, and keep them in their proper place. **I**

Mike Neal is CEO of [DecisionNext](#), a predictive analytics company for commodities industries with offices in San Francisco and Perth, Australia. Throughout his 25-year career, Neal has focused on taking empirical approaches to decision-making and using innovative science to solve supply chain challenges. In 2004, he co-founded SignalDemand with Dr. Hau Lee of Stanford University. He also co-founded DemandTec (now part of IBM) – the largest provider of consumer demand management software for retail price and promotion optimization. He holds an MBA from the J.B. Fuqua School of Business at Duke University and a bachelor's degree in economics and statistics from the University of Florida.

REFERENCES

1. <http://www.forbes.com/sites/gregsatell/2014/10/12/the-future-of-marketing-combines-big-data-with-human-intuition/2/>
2. <http://www.planadviser.com/Combined-Human-and-Robo-Advisers-Show-Promise/>
3. <https://books.google.com/books?id=e9yPQOoK0gcC&pg=PT9&lpg=PT9&dq=Steven+Cramton+and+Zackary+Stephen&source=bl&ots=xdVvKc0H1u&sig=WBKWUVN5cK9ONjl-bNcEKE8Z3l&hl=en&sa=X&ei=acSaVbe5M8uNyAT-ra7IBw&ved=0CE8Q6AEwCA#v=onepage&q=Steven%20Cramton%20and%20Zackary%20Stephen&f=false>
4. <http://www.retailon.com/predictive-analytics-for-retail>
5. <http://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/>
6. <https://practicalanalytics.wordpress.com/predictive-analytics-101/>
7. <https://hbr.org/2014/07/the-kind-of-work-humans-still-do-better-than-robots/>
8. <https://practicalanalytics.wordpress.com/predictive-analytics-101/>

My 'Philadelphia Story'

Poignant memories of a struggling young graduate student and the man who changed his career path.



BY VIJAY MEHROTRA

After a year's reprieve (fortunately for me, the INFORMS Annual Meeting was here in San Francisco last year), I find myself once again scrambling to get ready to leave my family and students behind for a few days in the middle of the semester to take a trip across the country to meet with my professional peers. This time around, the conference is in Philadelphia.

An operations research conference. In Philadelphia. Brings back a lot of very poignant and vivid memories for me.

Twenty-five years ago this fall, the ORSA/TIMS conference was held at the Wyndham Hotel in Philadelphia. In those days, I was a struggling young graduate student, plodding away without much confidence or direction and searching for a topic worthy of a Ph.D. thesis.

After reading some of Ward Whitt's papers on queueing networks, I came up with an idea that I thought might be a good dissertation topic. After that initial insight, however, I spent nearly a year looking at all sorts of things that led nowhere in particular.

I had a strong sense that I was on to something, and yet I lacked the gumption to simply pick up the

IMPOSSIBLE

Now you can switch
off impossible.

ODHeuristics is a brand new algorithm developed by Optimization Direct which achieves the previously impossible.

We created it for use with **CPLEX Optimization Studio** to solve difficult, large scale scheduling problems and it's much faster than your current solver.

So if your problems seem impossible or your solver is struggling to give good results, then our experts using **ODHeuristics** with **CPLEX** is the combination you've been waiting for.

Talk to us about your challenging optimization problem. **Find us at** optimizationdirect.com



**OPTIMIZATION
DIRECT**

phone and call the great Ward Whitt. I was in awe of this man – he had forgotten more about probability and queuing than I had ever known – and I was terrified of revealing my ignorance and my utter lack of genius. Heck, I didn't think he would even take my call.

When I happened upon a program of the 1990 ORSA/TIMS Conference in Philadelphia, I saw that Whitt would be presenting a paper. Although I had no thesis topic, no dissertation advisor and no conference travel funds, I decided in a desperate moment to go to the conference. No paper to present, no old friends to meet up with, no job interviews; I had no “real” business being at that conference. My sole purpose in getting on that airplane was to present my research idea to the one person that was surely knowledgeable enough to tell me whether it was a topic worth pursuing.

Traveling on Sunday from San Francisco, my flight seemed to take forever. I arrived in Philly late that night, and slept on a couch in my friend's mother's small suburban apartment. The next morning, I got up and caught a train into the center of the city. Once inside the conference hotel, I felt completely like a fish out of water. I had never been to an academic conference of any type before, and all of the logistics had only heightened my anxiety level. I sat through one unintelligible

presentation after another. I nervously roamed through the halls. All the while, I felt lonely and paralyzed, wondering how I might approach this wise man.

All day Monday, I couldn't find him. Translation: Consciously or unconsciously, it seemed that I had gone out of my way to avoid sessions where I might run into him. At the Monday night reception, I finally spotted him – thank God for name tags – but the loud and crowded room seemed like a less than ideal setting to talk about two-moment approximations. Translation: I chickened out.

My flight back to California was scheduled for Tuesday night, so my real last chance to corner him was a Tuesday morning session where he was presenting. Arriving late, I slipped into a chair in the back of the room, my heart pounding. Afterwards, I struggled to my feet and approached the small crowd that had gathered around the presenters. Meekly awaiting my opportunity to speak, I finally heard a pause in the conversation and, somewhat hesitantly, waded in.

“Do you perhaps know,” I asked timidly, “if anyone has done anything with the Fixed Population Mean [1] method?” He looked at me with pure kindness in his eyes, and his words were like music to my ears: “You know, no one really has, and I really think it is worth looking into further!”

CPLEX Optimization Studio®

Still the best optimizer and
modeler for the finance industry.

Now you can get it direct

Business
Partner



CPLEX Optimization Studio is well established as the leading, complete optimization software. For years it has proven effective in the finance industry for developing and deploying business models and optimizing business decisions.

Now there's a new way to get CPLEX – direct from the optimization industry experts.

Find out more at optimizationdirect.com



**OPTIMIZATION
DIRECT**

The IBM logo and the IBM Member Business Partner mark are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide.

*IBM ILOG CPLEX Optimization Studio is trademark of International Business Machines Corporation and used with permission.

Encouraged, I stammered out a few more sentences, trying hard not to humiliate myself by displaying my very serious intellectual limitations. Ward on the other hand treated me like a respected colleague, offering several enthusiastic off-the-cuff suggestions, as well as his Bell Labs business card. I thanked him profusely and repeatedly. Hell, I could have kissed him.

Back on campus, with a newfound sense of purpose, I spent the next few days writing up a 12-page document that outlined the problem statement and my proposed solution methodology. When I finished, I printed out a copy and sent it off to Ward.

Less than one week later, I received a thick envelope with Ward's initials and address on the return label. In it, I found detailed, line-by-line comments about the document that I had sent him (which ultimately became the introductory chapter of my thesis), along with several papers that were directly relevant to my research. This feedback did wonders to bolster my spirits, and the papers were an invaluable set of references for my work. I successfully finished my dissertation less than two years later.

No one conversation or experience determines the course of a lifetime. But finishing my Ph.D. – and perhaps even

more importantly, managing not to quit before I did, despite coming to the edge several times – did wonders for me. By managing to finish what I had started, I emerged from graduate school with a credential and a sense of confidence that together have turned out to be a gateway to the two very interesting careers, first as an analytics consultant and entrepreneur and more recently as a business school analytics professor.

Thinking back on this long-ago conference in Philadelphia, it was clearly a turning point for me. To this day, I feel deeply indebted to Ward Whitt, and I still marvel at the generosity of this extraordinary scientist and scholar with seemingly nothing to gain by helping out a young, nervous stranger like me.

May we all someday have the chance to experience such compassion. ■

Vijay Mehrotra (vmehrotra@usfca.edu) is a professor in the Department of Business Analytics and Information Systems at the University of San Francisco's School of Management and a longtime member of INFORMS.

Note: Portions of this column were previously published in <http://www.orms-today.org/orms-12-03/frsomething.html>

REFERENCES

1. <http://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1984.tb00084.x/abstract>



More than 17 million acres of farmland are lost to soil erosion every year. Many people who produce the world's food are living in poverty. Biodiversity is disappearing fast. And the challenge won't get any easier: by 2050, for example, 4 billion people will be living in countries with water scarcity.

THE CHALLENGE:

Using supplied data, determine which variety or varieties farmers should choose to plant next year (and in what proportions if multiple varieties are chosen).

CHALLENGE LAUNCH - Data Available Sunday, November 1, 2015

- Deadline for submissions January 15, 2016
- Selection of finalists March 4, 2016
- Finalist presentation (live or via telepresence)
at INFORMS Conference on Business Analytics and OR April 10–12, 2016



www.ideaconnection.com/syngenta-crop-challenge

ideaCONNECTION®

informs®

Healthcare and big data: Hype or unevenly distributed future?

Pundits predict healthcare will be the next frontier of big data analytics, but the future isn't quite here.



BY RAJIB GHOSH

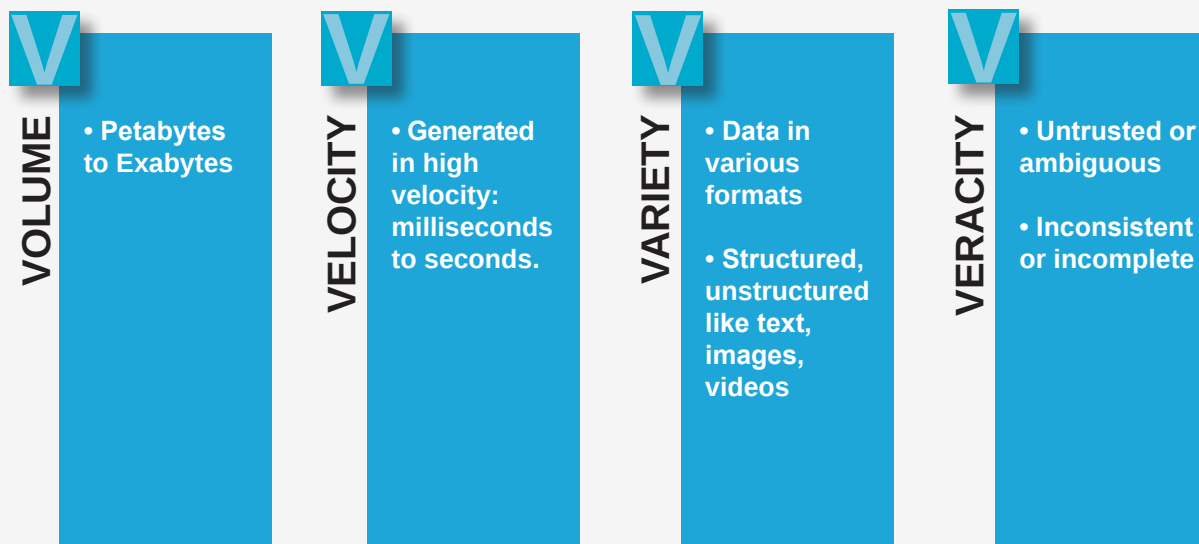
We had a quiet few months in the healthcare analytics industry news wise. Healthcare organizations, however, continued to build and deploy analytics solutions to optimize care, track outcomes and measure cost. Next month, Healthcare Information and Management Systems Society (HIMSS), a not-for-profit organization dedicated to improving healthcare quality, safety, cost and access through the use of information systems, will kick off a big data and healthcare analytics forum in Boston. I expect to see some interesting developments emerge out of that event.

As a precursor to the conference an interesting discussion has started to shape up: big data in healthcare. At the end of 2014, IDC [predicted](#) that by 2018, 50 percent of big data issues will become routine operational IT procedures. Technology companies and pundits predict healthcare will be the next frontier of big data analytics for a long time. In this article I will share my thoughts about the objectivity of this assessment.

IS HEALTHCARE DATA BIG DATA?

There are four attributes of a data set that makes it “big data”: volume, velocity, variety and veracity, which are described in Figure 1.

Figure 1: The 4 V's of Big Data



Do digitized healthcare data sets have those attributes? In most cases the answer is no. Digitization in healthcare happened not so long ago. There were some early adopters but not until the Affordable Care Act passed in 2010 – an incentive program created by the Office of the National Coordinator for Health IT for “meaningful use” of electronic health record (EHR) systems – adoption of EHR, the main tool for healthcare data digitization in provider organizations, was quite abysmal. In effect most healthcare delivery organizations started to digitize data only within the last five years or so. Some form of unstructured data like imaging data was stored in electronic format for quite some time. However, widespread

adoption lagged because of the high cost of such systems.

In a monolithic healthcare delivery model of the past, most of the digitized data were trapped in fragmented health IT systems that hardly interoperated. In other words, healthcare data volume never really approached the size of what we see in other big data domains such as social networking or consumer marketing.

Healthcare is very episodic in nature and therefore relatively low in volume and velocity. When a patient visits a doctor, a new encounter record gets created. Patient’s vital signs are recorded; allergies, symptoms and prescriptions are created. Once the episode is over, a

Healthcare data includes pharmacy data stored in the pharmacy system, structured claims data and physician text notes entered in EHR systems, and images stored in a picture archiving and communication system.

billing record is generated, and if the patient is insured a claim is sent to the clearinghouse for submission to the patient's insurance company. If a patient does not come back to see the doctor for the rest of the year or get admitted to a hospital for disease exacerbation, no more data for the patient gets added to any data set.

According to one [report](#), there were about 767,000 practicing doctors in the United States in 2013. [Another report](#) shows that on average, a doctor sees about 19 patients per day. This includes patients seen in an office, hospital or nursing home, on a house call or via an e-visit. This creates about 14.5 million encounter records per day assuming all episodes are unique, albeit that's not the case since we know [5 percent of the population](#) utilizes 50 percent of healthcare resources.

This is nothing compared to the data sets created per day on social media sites such as Facebook, Twitter, YouTube or Instagram and searches conducted by users on Google. Google does more than 3 billion searches per day based on [2012 data](#). On Facebook, users share [4.75 billion pieces of content](#) per day. In other words the healthcare data set does not grow exponentially at present compared to other big data sets out there.

Healthcare data includes but is not limited to pharmacy data stored in the pharmacy system, structured claims data and physician text notes entered in EHR systems, and images stored in a picture archiving and communication system. Veracity is the other key attribute of healthcare

data sets. In the absence of interoperability, many times data dictionaries used in various electronic systems are not consistent, which causes ambiguity.

As electronic data capture is a relatively new phenomenon in many healthcare organizations, data is likely not to be clean or complete. One may get a few data points per patient and only a subset of patients may have many such “little data” entries in the data set. Overall, this makes healthcare data quite unique and difficult for the purpose of training advance algorithms for predicting future events in patients.

WILL HEALTHCARE DATA BECOME BIG DATA?

So, is big data in healthcare another hype? Or is it the future that is already here but as economist William Gibson said, not evenly distributed?

Healthcare data volume can potentially explode if all the fitness or health monitoring wearable app data is added to a patient’s medical record or is combined in some other platform. That will increase data velocity substantially.

A recently published peer-reviewed [article](#) projects the growth of human genomics data to surpass the big data domains of YouTube, Twitter and astronomy by 2025 as the cost of genome sequencing falls rapidly and adoption

increases. However, scaling use of genomics to be used at the point of care in real time will take an exponential increase in the computational power and an order of magnitude in cost reduction. Along with that, many federal, state, local and HIPAA privacy regulations need to be resolved or rewritten before the data becomes available for widespread analytics.

So it is safe to say that healthcare is still quite far from being a big data domain. There is potential, however, that with the proliferation of the Internet of Things or connected devices during the next decade and advancement of genomics, healthcare data can become a big data domain. But in my opinion the future is just not here yet. We are still stuck in the present and that it is not that big. ■

Rajib Ghosh (rgghosh@hotmail.com) is an independent consultant and business advisor with 20 years of technology experience in various industry verticals where he had senior-level management roles in software engineering, program management, product management and business and strategy development. Ghosh spent a decade in the U.S. healthcare industry as part of a global ecosystem of medical device manufacturers, medical software companies and telehealth and telemedicine solution providers. He’s held senior positions at Hill-Rom, Solta Medical and Bosch Healthcare. His recent work interest includes public health and the field of IT-enabled sustainable healthcare delivery in the United States as well as emerging nations. Follow Ghosh on twitter [@ghosh_r](https://twitter.com/ghosh_r).

Agriculture analytics, new multimedia portal

The competition provides the chance to apply considerable talent in analytics to the terrible worldwide problem of hunger and poverty.

SYNGENTA, INFORMS OFFER CROP CHALLENGE

The Institute for Operations Research and the Management Sciences (INFORMS), the leading professional association in analytics and operations research, and [Syngenta](#), recently announced a new, joint award, the Syngenta Crop Challenge, which will focus on ways that analytics can address the problem of feeding millions of people throughout the world who face hunger every day.

“This new competition will give our members the chance to apply their considerable talent in analytics to the terrible worldwide problem of hunger and poverty,” says Glenn Wegryn, president of the Analytics Section of INFORMS and executive director of the Center for Business Analytics at the University of Cincinnati Carl H. Lindner College of Business.

The [Analytics Section](#) of INFORMS, comprising academics and practitioners, promotes the integration of a wide range of analytical techniques and supports activities that illuminate significant innovations and achievement in the growing field of analytics.

Nearly 7 million hectares of farmland are lost to soil erosion every year. Many people who produce the world’s food are living in poverty. Biodiversity is disappearing fast. And the challenge won’t get any easier; by 2050, for example, 4 billion people will be living in countries with water scarcity.



Syngenta supports farmers in the field with optimized seed trait, crop protection and seed treatment technology.

“At Syngenta, we bring plant potential to life by supporting farmers in the field with the very best seed trait, crop protection and seed treatment technology in industry,” says Joseph Byrum, Syngenta head of soybean seeds product development and leader of the Syngenta Crop Challenge. “Our commitment to innovation is unrivalled in the industry. We’re fundamentally transforming agricultural productivity through the use of operations research, as evidenced by our Edelman win, and we’re looking for exceptional talent to work with us through this challenge.”

Each year farmers have to make decisions about what crops to plant given uncertainties in expected weather conditions and knowledge about the soil at their respective farms. These decisions have important impacts; an unusual weather pattern can have disastrous impacts on crops, but planting to hedge against stressful weather patterns can dramatically reduce yields in normal years.

How can a farmer make seed variety decisions that optimally reduce risk and increase yield?

This is a question that experts in analytics and operations research can address by competing to win the new Syngenta Crop Challenge.

Case details will be announced at the 2015 [INFORMS Annual Meeting](#) in Philadelphia, which takes place Nov. 1-4. Teams must submit their report by January 2016. Finalists will be announced in March 2016. Finalists will make their presentation at the INFORMS Conference on Business Analytics and Operations Research in Orlando, Fla., set for April 10-12, 2016. Following the presentations, the winner of the inaugural Crop Challenge will be announced and will receive a \$5,000 prize.

For more information, visit: <http://www.ideaconnection.com/syngenta-crop-challenge/>.

Note: Syngenta won the 2015 [Franz Edelman Award](#) for achievement in operations research and the management sciences for its “Good Growth Through Advanced Analytics” program. The Edelman, considered the “Super Bowl” of advanced analytics application, was presented by INFORMS.

EDITOR'S CUT: NEW ONLINE MULTIMEDIA PORTAL

Leveraging the expertise of its operations research and analytics thought leaders, INFORMS Editor's Cut is a



comprehensive online multimedia portal designed to enrich the knowledge base of those interested in leading-edge analytics and operations research applied across a variety of topics. Each collection brings together key analytical insights and innovations from leading academics and practitioners on relevant topics through a multimedia experience. Introductory material and substantial research and industry contributions provide something from beginners who want the big picture to those who are looking for leading-edge research.

Editor's Cut is organized around topic areas that are new or rapidly evolving – where books or review articles likely haven't been developed. The bundle of multimedia materials is useful for a broad range of INFORMS members and a wider audience including the press. While effective for accessing the vast wealth of INFORMS content, the volumes go further to include publications, references and media from other sources, such as magazine articles or TED talks.

One goal of Editor's Cut is to anchor its subject matter area with a timely

INFORMS conference theme, providing a connection to ideas and presentations that represent work that may not be in print for months. For example, the inaugural volume, released July 27, is focused on [healthcare analytics](#) and was developed as part of the 2015 INFORMS Healthcare Conference held in Nashville. Beyond the journal and magazine articles, videos, TED talks and podcasts, Volume 1 also contains segments recorded at the conference including interviews with industry leaders.

The collection is made available in an open access format for all INFORMS members. As a member service, INFORMS members need not have subscriptions to all of the journals represented in the volume. Additionally, relevant conference attendees who are nonmembers will receive a time-limited token to access the entire volume. The general public will be able to access already openly available information as well as abstracts to academic articles. The goal is to provide a sophisticated yet simple-to-use platform for exploration. In addition to its compelling content and sleek user experience, what makes Editor's Cut a unique and powerful resource are the INFORMS experts leading and curating the collection.

Over the next months, Editor's Cut will continue to evolve, incorporating the social media platform INFORMS Connect



The inaugural volume is focused on healthcare analytics and was developed as part of the 2015 INFORMS Healthcare Conference.

for complementary dialogues on the various topics. New volumes of Editor's Cut will explore a range of ideas such as: sports, climate change/weather, big data, politics/social uprising/elections, humanitarianism, disaster preparedness/epidemics, information security/privacy, food/agriculture/farming and drones (delivery)/autonomous cars. |

— M. Eric Johnson and Anne Robinson

Driving curiosity as a culture in an analytics organization

Curiosity and continuous learning have become critical qualities for the analytics professional.



BY CHANDRAKANT
MAHESHWARI

During this decade, we have witnessed an unprecedented growth in information coupled with growth in technology, which has made access to information easy. This, in turn, has given rise to innovation and made the analytics business world spin faster than ever. Every time we turn around, we see a new technology or perspective coming our way.

As a result, curiosity and continuous learning have become critical qualities for the analytics professional, but the biggest challenge is knowing what to learn and where to explore. Even the brightest minds find the following challenges:

- Analytics is a moving target; with many options to explore but limited time and resources, deciding what to learn is crucial.
- Mastering the art of un-learning and then re-learning is a difficult task.
- For mid-level managers, it is important to understand the latest industry trends and innovation.
- For senior managers, it is important to understand what business impacts new innovations can bring.

- When it comes to learning, it's human nature to be impatient. While self-learning, professionals tend to jump or skip around key concepts. In the end, they do not properly understand the concepts and wind up unmotivated after losing their investment of time and resources.

These are the challenges faced by senior managers and leaders of organizations focused on business analytics. In a dynamic world where technology and ideas move and change with the blink of an eye, direct reports (DR) need to step up by helping their respective managers in determining the goals and required knowledge that are in line with the company's overall business objectives while keeping pace with the competition.

In appraisal interviews, the manager should ask his or her DR:

- What should I do to get my next promotion?
- How can you help me?
- What should I do to help you in helping me?

The first question is important in promoting innovative thinking. Whenever new ideas are discussed, the DR should be motivated to explore those ideas and innovate and hence develop a vision. Questions 2 and 3 will ensure that that innovative thought is implementable, measurable and aligned to the company's business objective.

Years back, the knowledge and understanding level in organizations was somewhat laddered. The more senior you were in the organization, the higher you were on the ladder, the more you knew. Managers generally knew and understood more

DISTRIBUTED THINKING

Recent growth in the analytics industry is attributed to the fact that organizations are able to implement distributed computing. If they are able to implement distributed thinking, there will be a paradigm shift in the industry. Driving curiosity as a culture is an approach to implement distributed thinking.

Successful managers and leaders will be those who would have mastered the art of listening to their direct reports.

about the company's business than their DRs. However, with the rise in freely available and easily accessible information, it is now a different ball game. The knowledge level and understanding of the business is more disperse. Today, many DRs may know and understand more than their managers when it comes to the company's business. In fact, progressive managers and senior leaders should make DRs responsible to learn more, and DRs should have the obligation to tell or teach their superiors what they do not know or understand.

There needs to be a paradigm shift in the attitude of senior management and leadership in the current dynamic scenario. Going forward, successful managers and leaders will be those who would have mastered the art of listening to their DRs, and successful DRs will be those who keep their eyes open for knowledge in this fast-moving market and then observe and assimilate whatever is relevant for their company's benefit. They need to up the ante in their maturity level by transferring the knowledge into implementable practical actions in their respective organizations.

This attitude is routinely followed by most companies at the CXO level. The difference between great organizations and others is that in great organizations this strategy is also followed at the most junior level. ■

Chandrakant Maheshwari (chandrakant721@yahoo.com) is a subject matter expert in the risk and regulatory practice at Genpact LLC. He leads consulting engagements with global financial institutions, advising clients on financial risk management. He is an avid blogger and blogs at <https://chandrakant721.wordpress.com/>.



Master of Science in
**ENGINEERING
MANAGEMENT**
Delivered 100% Online

A Technical Alternative to the MBA

Fast track option available – Finish in 12 months!

- 30-credit-hour curriculum with admission offered in Fall, Spring and Summer semesters
- Gain skills in demand by industrial, research and commercial firms
- Concentrations and graduate certificates available:
Logistics and Supply Chains,
Energy Systems, Lean Six Sigma and Systems Analytics

Now Accepting Applications
DistanceEd.uncc.edu
704-687-1281



UNC CHARLOTTE
Extended Academic Programs

Salary survey of predictive analytics professionals

Not surprisingly, compensation depends on experience, industry, job responsibilities and location.

According to a recently released salary survey of predictive analytics professionals (PAPs) conducted by Burtch Works, PAPs earn median annual salaries between \$76,000 and \$235,000, depending on, not surprisingly, their experience, industry, job responsibilities and location. Those numbers do not include bonuses, which the vast majority of PAPs receive and that can add significantly more income.

The survey, conducted over 12 months ending in April 2015, is based on responses from 1,757 PAPs who work for more than 800 different companies located across the United States.

Burtch Works, an executive recruiting firm focused on market research and analytics professionals, defines PAPs as those who can “apply sophisticated quantitative skills to data describing transactions, interactions or other behaviors of people to derive insights and prescribe actions.” Burtch Works distinguishes PAPs from business intelligence professionals and financial analysts by the enormous quantity of data with which they work, well beyond what can be managed in Excel. It should be noted that data scientists were excluded from the study because of



their distinguishing ability to work with unstructured data, resulting in different compensation.

The survey provides separate median salaries for three levels of “individual contributors” and “managers,” with Level 1 being the lowest compensation for each group based on experience, etc., and Level 3 the highest.

Among the findings:

- The median base salary of individual contributors varies from \$76,000 for those at Level 1 to \$125,000 for those at Level 3. The median base salary of managers varies from \$125,500 for those at Level 1 to \$235,000 for those at Level 3.

- The proportion of individual contributors eligible for a bonus varies from 69 percent for those at Level 1 to 87 percent for those at Level 3. More than 92 percent

of managers at all levels are eligible for bonuses. Regardless of job level, the median bonus paid to managers is significantly greater than the median bonus paid to individual contributors.

- Compensation of PAPs also varies based on characteristics including education level, location and industry of employment. Historically, PAPs working in the Northeast and on the West Coast have been paid more than other PAPs, and PAPs working for consulting firms were paid more than those working in other industries.

- A significant proportion of all PAPs are non-U.S. citizens. Among individual contributors at Levels 1 and 2, non-U.S. citizens are the majority.

- Entry-level individual contributors who hold a green card or are on an H-1B visa earn more than U.S. citizens. However, among Level 3 individual contributors, this trend reverses: The median base salary for those who hold an H-1B visa is 21.5 percent lower than for those who are U.S. citizens, and green card holders earn 7.7 percent less than U.S. citizens.

- Variation in PAPs’ compensation correlates most strongly with job category and with scope of responsibility (job level).

To download the complete report, [click here.](#) |

The reality of real time

Four important areas where real-time analytics can provide real value



BY LARRY SKOWRONEK



One of the most interesting trends in the analytics space today involves real-time analytics, and rightly so. The ability to analyze calls while they are in progress, and then provide the information the frontline staff can use to take immediate action, is a capability from which every contact center can benefit. The insights gleaned from a call while the customer is still on the line improves the agents and supervisors' ability to remain compliant with regulations and policies, increase the effectiveness of their sales efforts and improve the overall customer experience.

As the migration of voice systems from legacy TDM telephony to VoIP technology nears completion, the costs for deploying real-time audio analytics have fallen dramatically, bringing the application well within the reach of the broad population of contact centers of all sizes. Now that real-time analysis is a real possibility for companies in any industry, the question remains: How exactly can a company use real-time interaction analysis to achieve the best possible business results as it handles its customer interactions? Following are four key areas:



Real-time analytics is one of the most interesting trends in the analytics space today.

1. Identifying at-risk compliance situations

One of the most straightforward use cases for real-time voice analysis, applicable (to varying degrees) to all industries, is maintaining compliance, both with external regulations and internal policies and processes. By successfully managing the former, a business can sometimes avoid large fines from regulators, even as it protects its reputation in the marketplace. Just as important, tracking

internal compliance allows a company to work more efficiently and become far more capable of achieving its business objectives. How does real-time interaction analytics help companies manage both aspects of compliance?

Real-time analytics as an external compliance aid is especially important in industries that are heavily regulated, such as healthcare, financial services and debt recovery. Be it adhering to HIPAA, the Dodd Frank Act or ensuring that agents

are contacting the right party and delivering the mini-Miranda and recording disclosures, every regulated business has room for improvement. Where the goal is to maintain compliance with disclosure requirements, the real-time analytics system analyzes the call in progress, listening for the agents to state the required disclosures.

If a disclosure is not said in the required time frame, such as the first 90 seconds of the call, the system sends an alert to the agent, reminding them to make the disclosure. If they still don't deliver the disclosure in, say, the next 30 seconds, the system sends a reminder alert to the agent and an escalation alert to the supervisor. With this information, the agents have the best chance of saying the right thing at the right time.

In much the same way, reviewing calls as they are happening can give a company the ability to monitor how its agents are performing in real time to ensure that they are following company guidelines and providing the most effective and profitable service possible. For example, a company could provide alerts to agents when it detects that the agent is building a quote for a customer to remind them to disclose an activation fee (or warn them if it detects that they still have not mentioned the fee after

a sale). This disclosure very often will prevent the angry customer call back when they get their first bill. These little nudges have a powerful affect on agent performance.

Through real-time analytics, it is entirely possible for a company to monitor and manage potential compliance issues faster and more accurately than ever before.

2. Increase customer retention

For many businesses, increasing customer loyalty is their most important objective. Of course, any program designed to drive improvement in customer retention will have many components. One important retention strategy includes identifying areas of customer dissatisfaction by understanding customer sentiment, then taking pre-emptive action to improve it. Real-time analytics should employ both an acoustic and linguistic approach to the calculation of customer sentiment, with the ability to recognize pitch, tone, cross talk, laughter and, most importantly, the actual words and phrases spoken.

Based on these markers, the real-time system flags calls exhibiting negative sentiment and sends appropriate alerts to the agent desktop or supervisor console. These alerts offer instructions designed to save the call, such as

guidance on resolving the specific issue or a promotional offer the agent should make. The goal is to pinpoint an issue that can lead to customer defection at the moment it occurs and then address the issue before it's too late.

The science of predictive modeling is another approach to increasing customer retention. First, an analyst uses post-call analysis to identify interactions of customers known to have churned. The modeling process compares the topics discussed on the calls for those customers that eventually left with those discussed by the customers that stayed, finding the topics that lead to, or predict, that the customer will eventually churn. These results are then combined with additional data points such as customer spend and tenure data.

Based on this combined set of training data, the analytic process builds a predictive model that prioritizes each at-risk customer. When applied to a call in real time, this same churn prediction model is used to flag interactions with high propensity to churn scores and send alerts to agents and supervisors to take action right away. With this information, companies can implement real-time retention strategies such as special offers or discounts that can significantly minimize customer churn.

3. Increase operations and sales effectiveness

Similarly, the assistance provided by real-time analytics enables agents to handle calls with the confidence they will be far more effective. After post-call analysis identifies the behaviors and statements that are most effective at selling a product, overcoming objections, etc., the real-time system identifies the appropriate times to send alerts to the agents and supervisors, prompting them with specific up-sell, cross-sell and next-best offer queues. When used in such a fashion, real-time analytics has practical applications beyond understanding, designing and measuring sales tactics, including driving their improvement.

For debt recovery agencies, the use case is similar. The agency that used post-call analytics to understand the best practices for maximizing collections can then use that information to create real-time alerts that provide an agent with the best route to successful collection. As an agent handles a call, she receives alerts guiding her through those best practices to ensure that the proper strategies are being deployed to maximize the call's effectiveness. For example, the system might detect phrases such as "can't pay," "broke" or "not working." This could trigger an alert to the agent containing guidance on how to position a payment plan to which the customer can agree.

4. Keep real time in context with post-call analytics

Burdening agents with alerts or screen pops for every interesting situation that might occur defeats the purpose of any real-time monitoring program. Agents that receive more than just a few reminders will eventually ignore these alerts, suffering from “alert fatigue.” For real-time agent alerting to be successful, the business must alert agents only to the most important issues and actions, enabling them to handle their calls most effectively. To do this, management must know that the action requested by the real-time alert addresses an issue the business is facing on a large scale (more than just an individual agent basis) and that the requested action will have a meaningful impact.

Post-call analytics is the key to the successful prioritization of real-time alerts and requests for agent action. For instance, to optimize customer experience an analyst would examine the full range of call history, understanding which factors indicate negative customer sentiment and elevated stress and their correlation with low net promoter scores. Alerts would be sent to the agent offering techniques to prevent a potentially adverse outcome, including how to be empathetic and reminders to not talk over the customer.

Once the business knows how to change agent behavior for a negative customer situation, management can train everyone on the new ways to handle this type of scenario and configure the real-time alerts to support the agents with useful reminder information as they handle those tough customer calls.

CONCLUSION

As the volume of customer interactions across channels continues to grow, it is vitally important that companies not only take advantage of real-time analytics, but that they use the collected information to enact valuable changes. The key to getting the most from real time, as is true with any sort of analytics, is to take effective action on the findings. With each new insight discovered through post-call analysis, it is important to turn that information into the best practices that are then used during live calls and chats. It's with that effort that a company can discover just how valuable a tool real-time analytics can be. ■

Larry Skowronek is senior vice president of Product Management for [Nexidia](#), where he is responsible for the research and development of the overall market strategy, product strategy and development roadmap for Nexidia's speech analytics solutions. Skowronek has nearly two decades of experience working with contact center tools and management. He has led product management, quality assurance and consulting organizations for contact center management software vendors across the industry.

Become More.



EARN AN AACSB-ACCREDITED MBA FROM THE UNIVERSITY OF SOUTH DAKOTA

Advance your career with an online Master of Business Administration with a specialization in Business Analytics through the Beacom School of Business at the University of South Dakota.

Solve key business problems using big data in areas of:

- ✓ Sales
- ✓ Marketing
- ✓ Operations Management
- ✓ Finance
- ✓ Economics
- ✓ Strategy

Grad Performance in to 3% -

Our students score in the 97th percentile in a national exam for graduate business students.

Real Professors -

Learn from the same faculty who teach on campus.

Be Ready to Lead -

Our programs emphasize decision making, problem solving and social responsibility.

Flexible Schedules -

Work full-time while earning your advanced degree.
Our courses are built around manageable time commitments.

**DIVISION OF CONTINUING
& DISTANCE EDUCATION**



UNIVERSITY OF
SOUTH DAKOTA

414 E. Clark St. | Vermillion, SD 57069
800-233-7937 | 605-658-6140
grad@usd.edu

Learn more: www.usd.edu/onlinemba



The University of South Dakota's Beacom School of Business has been accredited since 1949. USD's M.B.A. Program has been accredited by AACSB since 1965.

Big data in marketing analytics



BY MOUSUMI GHOSH

Big data is the biggest game-changing opportunity and paradigm shift for marketing since the invention of the phone or the Internet going mainstream. Big data refers to the ever-increasing volume, velocity, variety, variability and complexity of information. For marketing organizations, big data is the fundamental consequence of the new marketing landscape, born from the digital world we now live in. The term “big data” doesn’t just refer to the data itself; it also refers to the challenges, capabilities and competencies associated with storing and analyzing such huge data sets to support a level

of decision-making that is more accurate and timely than anything previously attempted: big data-driven decision-making.

Organizations today face overwhelming amounts of data, organizational complexity, rapidly changing customer behaviors and increased competitive pressures. New technologies, as well as rapidly proliferating channels and platforms, have created a massively complex environment. Data worldwide is growing 40 percent per year, a rate of growth that is daunting for any marketing and sales leader. Many marketers may feel like data has always been big – and in some ways, it has. But think about the customer data businesses collected



Having big data doesn't automatically lead to better marketing.

20 years ago – point of sale transaction data, responses to direct mail campaigns, coupon redemption, etc. Then about the customer data collected today – online purchase data, click-through rates, browsing behavior, social media interactions, mobile device usage, geolocation data, etc. Comparatively speaking, there's no comparison.

Having big data doesn't automatically lead to better marketing. We can think of big data as a secret ingredient, raw material and an essential element. It's not the data itself that's so important. Rather, it's the insights derived from big data, the decisions we make and the actions we take that make all the difference.

Three types of big data are key for marketing:

1. Customer: The big data category most familiar to marketing may include behavioral, attitudinal and transactional metrics from such sources as marketing campaigns, points of sale, websites, customer surveys, social media, online communities and loyalty programs.

2. Operational: This big data category typically includes objective metrics that measure the quality of marketing processes relating to marketing operations, resource allocation, asset management, budgetary controls, etc.

3. Financial: Typically housed in an organization's financial systems, this big data category may include sales, revenue, profits and other objective data types that measure the financial health of the organization.

Organizations that want to succeed in marketing should do the following things well:

1. Successful discovery of new opportunities. Successful discovery requires building a data advantage by pulling in relevant data sets from both within and outside the company. Relying on mass analysis of those data, however, is often a recipe for failure. Analytics leaders need to go beyond broad goals such as "increase wallet share" and get down to a level of specificity that is meaningful. They need to use digital information to better target buyers and use heaps of analytics to learn more about target buyers than ever known before. Modern marketers should shed light on a more granular level of detail, such as: which websites a user frequents most often, which social media profiles they have and use, and even which buttons they click on a given website. The "ideal customer profiles" can easily be targeted with the big data.

2. Understand consumer decision journey. Today's channel-surfing consumer is comfortable using an array of devices,

tools and technologies to fulfill a task. Understanding that decision journey is critical for identifying battlegrounds to either win new customers or keep existing ones from defecting to competitors. Marketing and sales leaders need to develop complete pictures of their customers so they can create messages and products that are relevant to them.

By combining big data with an integrated marketing management strategy, marketing organizations can make a substantial impact in these key areas:

- Customer engagement: Big data can deliver insight into not just who your customers are, but where they are, what they want, how they want to be contacted and when.
- Customer retention and loyalty: Big data can help you discover what influences customer loyalty and what keeps them coming back again and again.
- Marketing optimization/performance: With big data, you can determine the optimal marketing spend across multiple channels, as well as continuously optimize marketing programs through testing, measurement and analysis.

3. Monitor Google Trends to inform your global/local strategy. Google Trends is probably the most approachable method of utilizing big data. Google Trends showcases trending topics by quantifying

how often a particular search-term is entered relative to the total search-volume. Global marketers can use Google Trends to assess the popularity of certain topics across countries, languages or other constituencies they might be interested in, or stay informed on what topics are cool, hip, top-of-mind or relevant to their buyers.

4. Create real-time personalization to buyers. Marketers need to send the right message at the right time.

Timeliness and relevancy aren't just qualities of the Fourth Estate; they're also the foundation of successful marketing campaigns, e-mail click-through rates and consumer engagement with your brand.

Big data gives marketers timely insights into who is interested or engaging with their product or content in real time. Tying buyer digital behavior into your CRM systems and marketing automation software allows you to track the topics that your buyers are most interested in and send them content that makes the most sense to develop those ideas or extrapolate on those topics.

5. Identify the specific content that moves buyers down the sales funnel.

How successful was a singular blog or social post at generating revenue? Before big data that was an unanswerable question. We executed on social media strategies and content creation because

we had a feeling that it was working, but we had no way to back that claim. Now, marketers can distill the effectiveness of a marketing push down to tweet. Tools like content scoring illuminate which individual content assets were successful to a closed/won deal, and which were inefficient. This allows marketers to hone the strategies around the content topics or types that resonate with their buyers the most, and truly compel them to purchase.

6. Make it quick and simple. Companies need to invest in an automated "algorithmic marketing," an approach that allows for the processing of vast amounts of data through a "self-learning" process to create better and more relevant interactions with consumers. That can include predictive statistics, machine learning and natural language mining. These systems can track key words automatically, for example, and make updates every few seconds based on changing search terms used, ad costs or customer behavior. It can make price changes on the fly across thousands of products based on customer preference, price comparisons, inventory and predictive analysis. ■

Mousumi Ghosh (mousumi.ghosh@chase.com) is a vice president and senior manager of pricing analytics at JPMorgan Chase & Co. where she works with regional sales, credit teams and senior management. She is a member of INFORMS.

Simulated worlds

Driven by questions, fueled by thought and realized by simulation



BY JAMES J. SWAIN

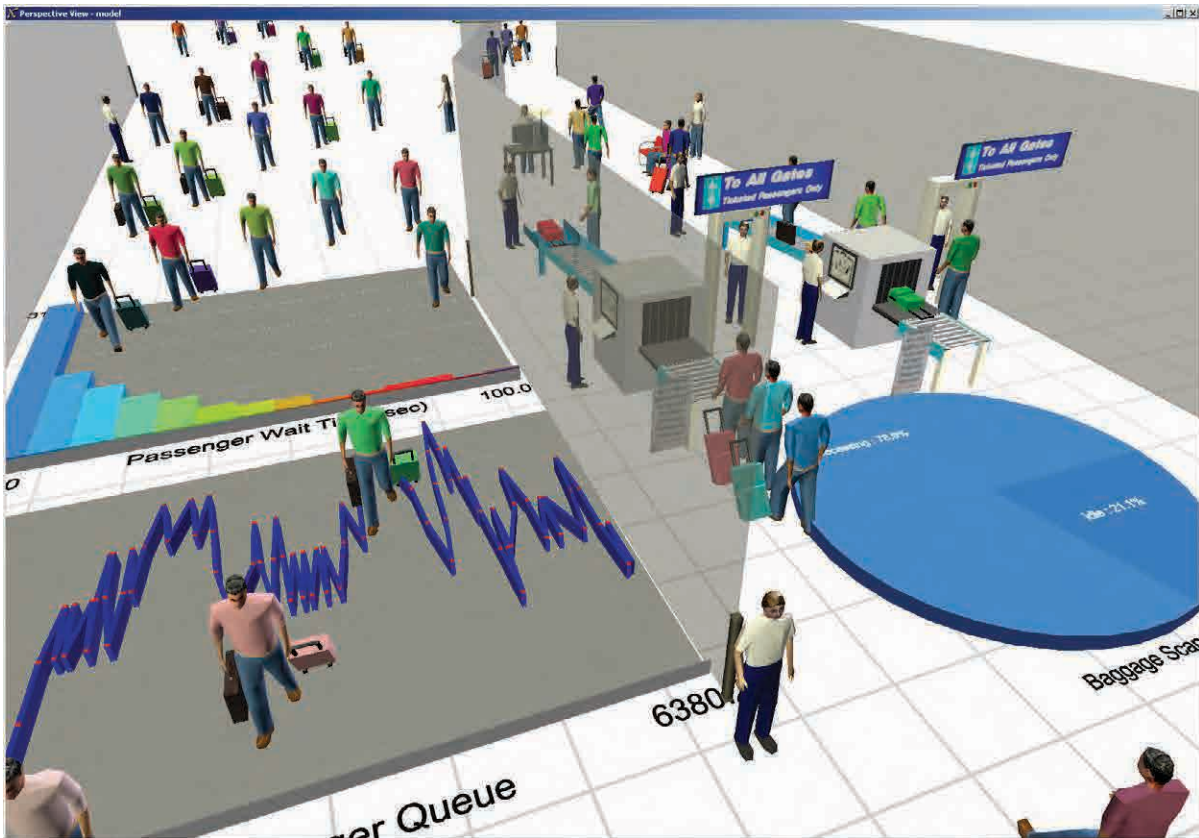


Our world and that of our clients is a complex place and fraught with uncertainty. The rewards for innovation and improved efficiency are immense if they can be realized, but there is risk as well: Can the innovations be implemented amid the connections that tie our system to our partners and customers? To gain knowledge, we turn to analysis and to experimentation. Yet how can experiments be run when the world will not stand still to be questioned?

For more than a half century, simulation has provided a major and constantly evolving tool to stand in proxy to the world, to provide a test bed for ideas and a basis for both experimentation and understanding. During that time, modeling has gone

from an exotic tool that was limited to a few research centers to widespread, almost ubiquitous use. Simulation-based gaming has become a major presence in entertainment, and the size of that market has propelled graphic displays and the hardware that makes it routinely available. It is surely no longer necessary to explain what simulation is or convince clients to accept its results; they (or their children) are already familiar with the world or deeply immersed within it.

Computer-based simulation is an immense field that spans models that explore the collisions of galaxies, the flow of turbulence over a turbine blade, or examines the interactions among subatomic particles. Our focus in this survey is restricted to the realm of discrete-event



Airport example: Simulation provides a tool to stand in proxy to the real world.

simulation models that are particularly suited to the operation of parts, patients, vehicles and such in widely divergent fields of manufacturing, healthcare and other services settings, logistics, transportation or military operations to name a few areas of application.

This 10th biennial simulation survey provides a snapshot of a vital and robust area of analysis software. The vendors document both the growth and sophistication of their tools. Of perhaps greater interest is to examine the case studies and the white papers that many provide,

demonstrating how the tools were successfully applied and the benefits that were obtained from the effort.

WHY IS SIMULATION SO SUCCESSFUL?

Simulation has grown as a tool because it provides us with a cost-effective and immensely powerful tool for exploration that is limited only by imagination. It is a tool that is only possible because of the computer, and its continued growth is made possible by maturation in diverse fields including modeling, analysis, computing capacity and speed, and programming tools.

Beside the growth in simulation tools there has been a steady growth in the body of knowledge about simulation and its application. Data is more accessible and more readily incorporated into models, thus increasing the level of detail that can be represented. And, of course, simulation tools incorporate the lessons of the experience in the field. Meanwhile, enabling technologies, such as random number generation, have improved over the decades. The state of the art in random number generation provides virtually unlimited, high-quality random numbers.

Another reason that simulation is so successful is that experimentation with the real world is rarely feasible, and there are economic and practical limitations to range and amount of experimentation that is possible. The statistician George E. P. Box once observed that “all models are wrong, but some are useful,” and this applies to simulation models as it did with statistical approximations. Simulation has been successful because it is possible to build simulation models that are sufficiently valid for an analysis to be useful.

ADVANTAGES OF SIMULATION

One of the primary advantages of simulation is that it is a constructive tool. Simulation models attempt to reproduce the states and trajectories of the actual system, using transformations, rules or

SURVEY DIRECTORY & DATA

To view the directory of simulation software vendors, software products and survey results, [click here](#).

procedures that the real system would use. This has many advantages. First, it means that model building is based upon a description of the process that is already available. Moreover, many modeling languages have evolved constructs such as servers or workstations that fit many situations, or have evolved specialized instances applicable to particular fields such as manufacturing or healthcare, with both logic and animation built in.

Constructive models are certainly easier to understand and to explain. While a queueing formula may provide a prediction of the mean waiting time in a queue, its derivation is opaque to most clients and its application is further limited. Of course, the simulation model provides an estimate, but in many cases increases in computing power and speed reduce the delay in obtaining accurate estimates, and the price is generally well worth the gain in understandability and range of application.

In statistical modeling, a mechanistic model is one that arises from the theory of the process, often as a solution to a set

of differential equations, as is the case in chemical models. Box observed that mechanistic models had several advantages over empirical regression models. Such models, he noted, contributed to our knowledge of the phenomenon, provided a better basis for extrapolation, and tended to require fewer parameters (“parsimony”) while providing a better estimate of the output. Simulation models, by their emulation of the process being simulated, seem to be a type of mechanistic model.

This is critical if simulation models are going to be used as the basis of experimentation or optimization where a degree of extrapolation is necessary.

Of course, a critical advantage of constructive models is that they generate the states of the system as they evolve, driving animation and allowing a detailed analysis, including detailed statistical observations about the system. The former provides a powerful starting point for both validation and user credence.

Share Your Model Online. Simply.



Join more than 360,000 students and professionals using Forio's platform to share insights and analysis.

Build your own or collaborate with us.

/forio

Discover the Potential | forio.com/informs

Providing a visually accurate portrayal of the system in question gives increased confidence in any analysis derived from the model. It can also be used to give snapshots about specific examples of typical or worst case behavior that can be examined in detail.

Animation can also be useful as a tool to facilitate communication and understanding among team members, who often come from different disciplines and backgrounds.

Many of the tools in this survey provide libraries of images that allow the realistic portrayal of equipment, parts or people to make the visualization accurate. Most now provide 3-D animation, which can be examined from different viewpoints and at various levels of detail. Here simulation benefits from the great strides made in the gaming industry for representation and perhaps the growth of graphics processors on home computers for gamers.

Improvements in programming have allowed some models to incorporate schedules, to link to scheduling algorithms or to run historical inputs (data-driven simulation). With some it is possible to make decisions based upon the state of the system, as the real system might do. For automation systems, some provide emulation of the control systems that are to be used to validate their operation in the field.

SIMULATION, ANALYTICS & OPTIMIZATION

It has long been appreciated that the constructive approach allows a more detailed analysis of the system statistics. Whereas the queueing formula may be limited to a prediction of the mean, in the simulation model it is possible to observe the variability of the response, not to mention relations among different statistics. This is just the beginning of the possibilities. Increasingly it has been noted that analytics and simulation might be good partners. Simulation can provide detailed data that can now be stored and accessed for insights using the tools of analytics, both within simulation replications and between different scenarios or cases. The 2015 Winter Simulation Conference has this connection as its theme: "Exploring Big Data through Simulation."

Once a valid and credible model has been built, it can be used for experimentation and optimization. In the former case, many of the simulation tools have the capability of generating experimental results automatically. They are sometimes called scenario generators or experiments. That is, the user can specify the parameters values to be run and the complete set of replications can be performed automatically, with the results available as a group. The results can then be compared graphically and statistically.

In other cases, the results can be exported for a more detailed analysis using statistical software.

Optimization is an important use of simulation. Optimizers such as OptQuest usually use heuristic approaches (e.g., Tabu search) that seem to work well where assumptions about the simulation responses are more limited. Optimization is used when the possible scenarios are too numerous to compare directly, so an algorithm is used to search among the parameter values to obtain improved responses.

Several products have the ability to perform sensitivity analysis on a model scenario given uncertainty in the model parameters. This can be used to determine the parameters that the responses are most sensitive to.

A final reason for simulation success is that the number of people trained in simulation has steadily increased. The Rockwell Software (Arena) website notes that 25,000 students take courses in the Arena software annually. Many of the simulation software in the survey provide low-cost student versions of their software and have textbooks that can be used for an introduction to their software. All industrial

Winter Simulation Conference 2015

The Winter Simulation Conference (WSC) has been the premier international forum for disseminating recent advances in the field of systems simulation for more than 40 years. The longest-running conference devoted to simulation as a discipline, this year's WSC will be held Dec. 6-9 in Huntington Beach, Calif., at the Hyatt Regency Huntington Beach. The theme for WSC 2015 is "Social and Behavioral Simulation."

Simulation is widely applied across a diverse range of endeavors. This is reflected in WSC 2015, which includes tracks devoted to manufacturing, health-care, logistics, gaming, military applications, big data and decision-making and more. WSC offers an invaluable educational opportunity for both novices and experts alike, with a large segment of each program devoted to introductory and advanced tutorials. These are carefully designed to address the needs of simulation professionals at all levels of expertise and are presented by prominent individuals in the field.

Vendor workshops, software tutorials and exhibits from leading software vendors are also featured at WSC 2015.

Register now at www.wintersim.org/2015/.

The range and variety of simulation continues to grow, reflecting the robustness of the products and increasing sophistication of users.

engineering programs include simulation in their curriculum, and simulation is also offered in many business programs as well.

THE SURVEY

This survey is the 10th biennial survey of simulation software for discrete-event systems simulation and related products (Swain, 2013 [1]). All product information has been provided by the vendors. Products that run on personal computers to perform discrete-event simulation have been emphasized, since these are the most suitable for usage in management science and operations research. Simulation products whose primary capability is continuous simulation (systems of differential equations observed in physical systems) or training (e.g., aircraft simulators) are omitted here.

There are 55 products listed in the survey, taken from 31 vendors who submitted for the survey, once again surpassing the last survey. The range and variety of these products continues to grow, reflecting the robustness of the products and the increasing sophistication of the users. The information elicited in the survey from the vendors is intended to provide a general gauge of the product's capability, special features and usage. This survey includes information about experimental run control (e.g., experimental design and automated scenario run capabilities) and special viewing features, including the ability to produce animations or demonstrations that can run independent of the simulation software itself. A separate listing gives contact information for all of the vendors whose products are in the survey. This survey

Which organization has the best O.R. department in the world?



call for nominations

DEADLINE TO APPLY IS DECEMBER 1, 2015

The Institute for Operations Research and the Management Sciences annually awards the INFORMS Prize for effective integration of Operations Research/Management Science (OR/MS) and advanced analytics into organizational decision making. The award is given to an organization that has repeatedly applied the principles of OR/MS and advanced analytics in pioneering, varied, novel, and lasting ways.

log onto: www.informs.org/informsprize for more information
or contact Julia Morrison, Lead O.R. Scientist
voice: +1 301-380-3067
email: julia.e.morrison@marriott.com



2016

Tell us why the title should be yours.



informatics

Analytics Professionals:
Regardless of where you work, let INFORMS help you connect with others in your field.

twitter

LinkedIn

facebook

Blogs

www.informs.org/Connect-with-People/Social-Networking

is available online (<http://www.orms-today.org/surveys/Simulation/Simulation.html>) and will include vendors who missed the publishing deadline. Of course, most of the vendors provide their own websites with further details about their products. Many of the vendors also have active users groups that share experience in the specialized use of the software and are provided with special access to training and program updates.

There are a number of technical and professional organizations and conferences devoted to the application and methodology of simulation. The INFORMS publications Management Science, Operations Research and Interfaces publish articles on simulation. The INFORMS Simulation Society sponsors simulation sessions at the national INFORMS meeting and makes awards for both the best simulation publication and recognition of service in the area, including the Lifetime Achievement Award for service to the area of simulation. For further information about the Simulation Society, visit: <https://www.informs.org/Community/Simulation-Society/>. This site also contains links to many vendors of simulation products and sources of information about simulation, simulation education and references about simulation.

The Society for Modeling and Simulation International (www.scs.org), also devoted

to all aspects of simulation, holds conferences in the spring, summer and fall that cover all aspects of simulation practice and theory. The AlaSim International Conference is a relatively new conference hosted in Huntsville, Ala., with a focus on DoD and other government applications of simulation. The next AlaSim conference is scheduled for May 2016.

The INFORMS Simulation Society and the Society for Modeling and Simulation are both sponsors of the annual Winter Simulation Conference. This year's conference will be held Dec. 6-9 in Huntington Beach, Calif. As in past years the conference will be held together with the Modeling and Analysis of Semiconductor Manufacturing (MASM) conference. Further information and registration information is available from the site www.wintersim.org. This site also links to the complete contents of the Proceedings of the Winter Simulation Conference from 1968 to 2012 for ready access to research and applications of simulation. ■

James J. Swain (swainjj@uah.edu) is a professor in the ISEEM department at the University of Alabama in Huntsville. He is a member of INFORMS.

REFERENCES

1. Swain, J. J., "Discrete Event Simulation Software Tools: A better reality," *OR/MS Today*, Vol. 40, No. 5, October 2013, pp. 48--59.



THE NATION'S FIRST
Associate in Applied Science (A.A.S.)
degree in Business Analytics
on campus or online.



Why Study Business Analytics?

The Business Analytics curriculum is designed to provide students with the knowledge and the skills necessary for employment and growth in analytical professions. Business Analysts process and analyze essential information about business operations and also assimilate data for forecasting purposes. Students will complete course work in business analytics, including general theory, best practices, data mining, data warehousing, predictive modeling, project operations management, statistical analysis, and software packages. Related skills include business communication, critical thinking and decision making. The curriculum is hands-on, with an emphasis on application of theoretical and practical concepts. Students will engage with the latest tools and technology utilized in today's analytics fields.

Accelerated Executive Program

Our accelerated learning options allow students to complete certificate credentials in two semesters part time or one semester full time. Accelerated options are available for the Business Intelligence and the Business Analyst certificates.

Questions?

Tanya Scott
Director, Business Analytics
919-866-7106
tescott1@waketech.edu

Credential options

- Enroll in one or several:
- AAS degree
- Certificates: Business Intelligence, Business Analyst, Finance Analytics, Marketing Analytics, and Logistics Analytics

Flexibility

- Open-door enrollment
- Courses are offered in the fall and spring
- Courses can be taken online or on campus
- Competitively priced tuition

Gain skills in:

- Data gathering
- Collating
- Cleaning
- Statistical Modeling
- Visualization
- Analysis
- Reporting
- Decision making
- Presentation

Use data and analysis tools:

- Advanced Excel
- Tableau
- Analytics Programming
- SAS Enterprise Guide
- SAS Enterprise Miner
- SPSS Modeler
- MicroStrategy

Funded in full by a \$2.9 million Dept. of Labor Trade Adjustment Assistance Community College & Career Training (DOLTAACCCT) grant.



businessanalytics.waketech.edu



Mother Nature meets the Internet of Things



BY ATANU BASU (PICTURED) AND GABE SANTOS

W

hat does the oil and gas industry have in common with agriculture? More than you think. Both industries are influenced by Mother Nature in complex ways that we understand better every day. Advancements in computational technologies are making it possible to interpret information about nature's impact

on these industries in ways not possible just a few years ago.

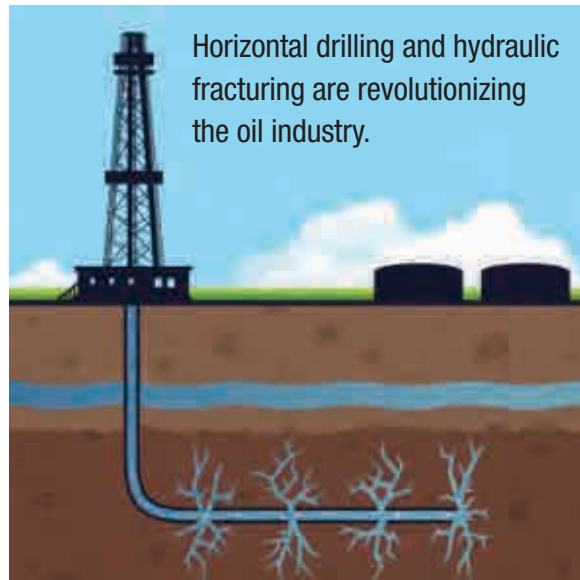
It used to be all about understanding your numbers and then making a decision based on that structured data. But now industries, including complex ones like oil and gas, are using sensors to tap into information such as sounds, images, videos and text. That unstructured

information can be mined just like numbers to unearth valuable insights that would not have been known otherwise.

In oil and gas, Mother Nature has packed several surprises and challenges hidden underground. She put oil inside rocks deep beneath the land, and though the industry has known this for a long time, it didn't know how to extract the oil economically. Horizontal drilling and hydraulic fracturing are revolutionizing the industry and helped catapult the United States into the world's No. 1 producer of oil and gas [1]. But even with these technologies, a lot of oil is being missed. That is where advanced sensors capturing data, i.e., the Internet of Things, is augmenting production capabilities.

For example, fiber optic sensors laid along a horizontal well for miles underground record and report – every foot and every second – the sound, pressure, temperature and more of hydraulic fracturing and energy production. Well operators, equipped with this information, get a new understanding of how a well is performing over time. In other words, the sensors help explain what is happening to Mother Nature during the energy extraction process.

With these disparate sources of information flooding in to operators, they need a way to connect the dots and take the best actions possible for each well. That next



step is the province of prescriptive analytics, a technology that is domain agnostic by design. Prescriptive analytics software takes all of that structured and unstructured data and predicts what's ahead and prescribes what can be done to achieve the best possible outcome. Working like a central nervous system, prescriptive analytics takes all of that data through numerous mathematical and computational algorithms and creates a recipe that tells operators how to change their settings for such things as water pressure and chemicals to get more oil before moving on to the next well.

THE AGRICULTURE ANGLE

What does all of this have to do with agriculture? The same principles used to maximize profits in the oil and

gas industry can be a boon for farming. Like in oil and gas, sensing equipment is greatly empowering farmers by providing real-time analysis of crop condition, crop stress, growth stages, disease pressure, insect pressure and problem areas within fields.

Precision agriculture started with the development and commercial availability of GPS and satellite imagery and allowed for more accurate production areas, soil mapping and soil sampling. As a result, producers, agronomists and other experts could pinpoint the exact location and scope of problems in their fields. But field and yield information is only valuable to farmers if it informs a management decision or agronomic practice.

With the application of prescriptive analytics, farmers are not entirely dependent on nature's whims. They now have a better way to use the information they've derived from a broad spectrum of structured (i.e., soil pH levels) and unstructured data sources (i.e., aerial satellite imagery).

Variable rate application has enabled farmers to target nutrients where they are most needed, rather than blindly broadcasting crop nutrients across the entire farm. Moreover, agronomists are now able to make specific prescriptions for growers based on their goals, and these prescriptions can be made to maximize yields, build and maintain nutrients over time, and reduce costs.

Agriculture represents the next frontier for prescriptive analytics. Precision agriculture is giving way to decision agriculture. It's a prescription for happy farmers. ■

Atanu Basu is CEO & president of Ayata, an Austin, Texas-based company that develops prescriptive analytics software solutions. He is a member of INFORMS. Gabe Santos is managing partner of Homestead Capital, a private investment partnership formed exclusively to invest in operating farmland. A version of this article appeared in Global AgInvesting News. Reprinted with permission.

REFERENCES

1. <http://www.eia.gov/todayinenergy/detail.cfm?id=20692>

Join the Analytics Section of INFORMS

For more information, visit:

<http://www.informs.org/Community/Analytics/Membership>





**Is the largest association for analytics in the
center of your professional network?**

It should be.

- INFORMS Allows You to Network With Your Professional Peers and Others Who Share Your Interests
- INFORMS Connect, the New Member-only, Online Community Lets You Network with your colleagues
- INFORMS Provides Unsurpassed Networking Opportunities Available in INFORMS Communities and at Meetings
- INFORMS Offers Certification for Analytics Professionals
- INFORMS Helps You Take Leadership Roles to Help Build your Professional Profile
- INFORMS Career Center Provides You with the Industry's Leading Job Board

Join Online Today!
<http://join.informs.org>

informs[®]

Data reduction for everyone



BY WILL TOWLER

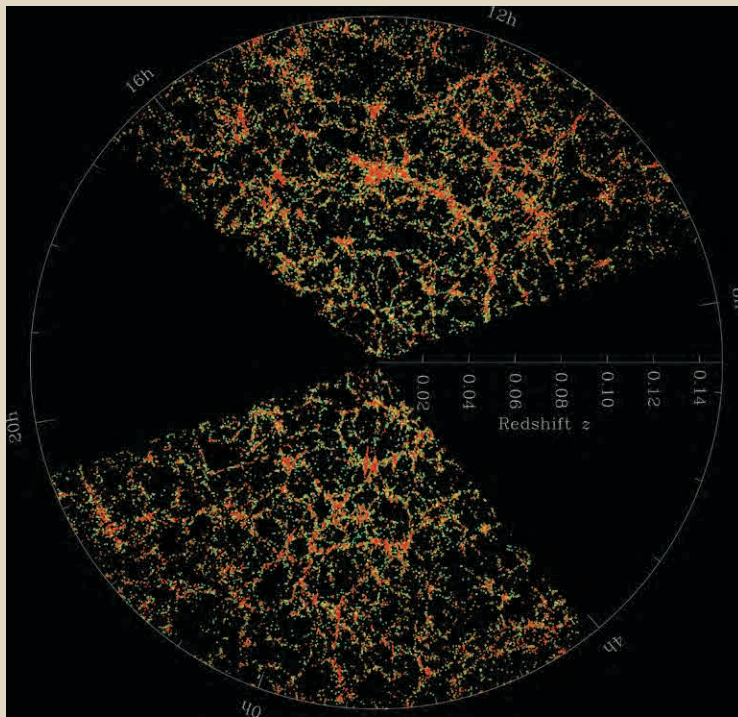
Data reduction, or the distillation of multitudinous data into meaningful parts, is playing an increasingly important role in understanding the world around us and improving the quality of life. Consider the following examples:

- Classifying space survey results to better define the structure of the universe and its evolution (see sidebar)
- Analyzing patient biological data and health records for improved diagnosis and treatment programs (see sidebar)
- Segmenting consumers based on their behavior and attitudes to enhance marketing efficiencies

- Better understanding the nature of criminal activities to improve counter-measures
- Optimizing mineral exploration efforts (see sidebar) and decoding genealogy

PROLIFERATING TECHNIQUES

The large number of use cases have led to a proliferation in data reduction techniques, many of which fall under the umbrellas of data classification and dimension reduction. Wikipedia claims there could be more than 100 published clustering algorithms used to classify data. Even more common techniques such as K-means and Hierarchical



Mapping the Universe

Cluster analysis is used to classify galaxies based on spectral data collected through initiatives such as the Sloan Digital Sky Survey (SDSS). Greater insight into the structure of the universe is helping us to better understand its history and future.

Source: M. Blanton and [SDSS](http://www.sdss.org) collaboration, www.sdss.org

Cluster Analysis can vary in execution (e.g., differing in stat tests, distance measures and linkage criteria).

There's a seemingly equal slew of dimension-reduction methods. And again, even the more popular techniques such as principal component analysis and factor analysis can vary in execution (e.g., differing in extraction and axis rotation methods).

EXPLORING DIFFERENCES

Deciding on the most appropriate data reduction technique for a given situation can be daunting, making it tempting to simply rely on the familiar. However, different methods can produce significantly

different results. To make the case we compare K-means and agglomerative hierarchical clustering (AHC) to segment countries based on export composition. We also compare different configurations of principal component analysis (PCA) to determine underlying discriminators. Data are drawn from the OECD Structural Analysis (STAN) Database and are analyzed with XLSTAT, an easy-to-use Excel add-in.

Results from K-means and AHC differ noticeably in statistical efficiency and group memberships. AHC (Euclidean distance and Ward's linkage) achieves 80 percent explained variance with 25 clusters, while K-means

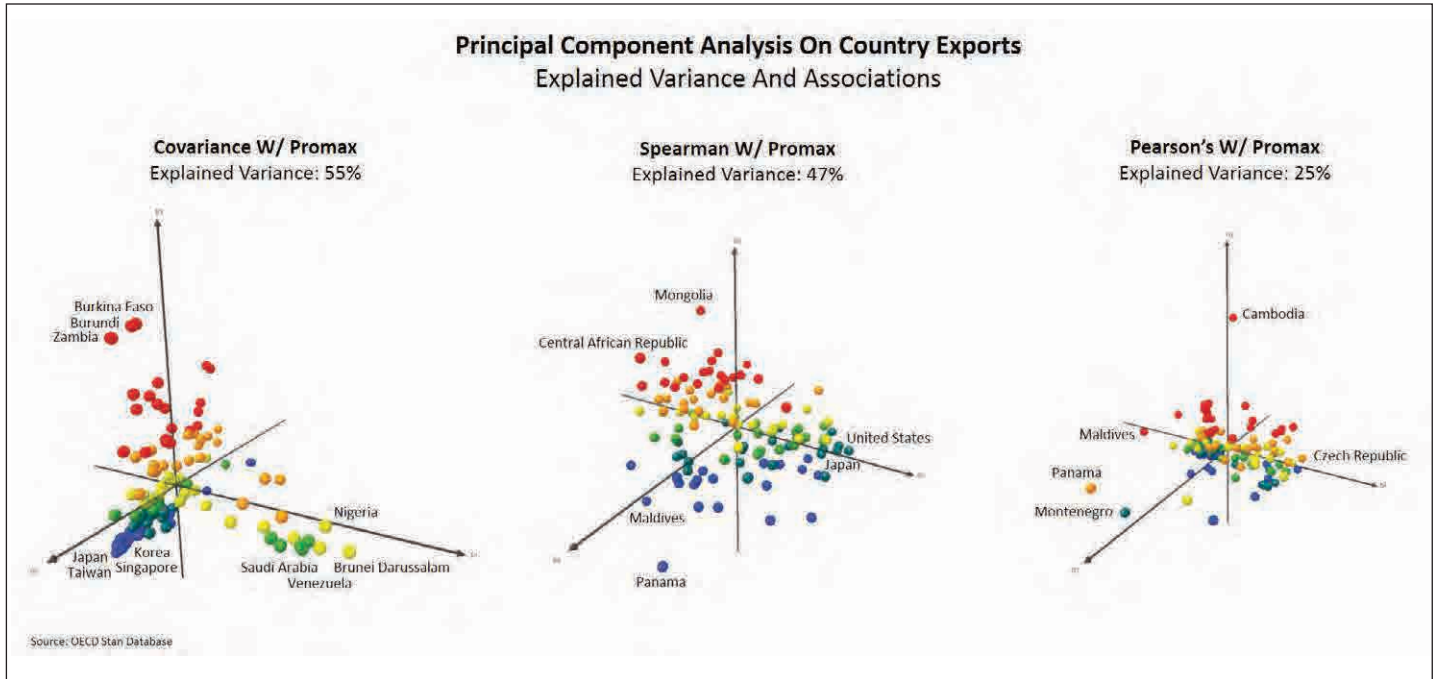


Figure 1: Principal component analysis on country exports.

(Trace stat test) requires 30 clusters to achieve the same explained variance. Other K-means and AHC configurations are less efficient. As for segment characteristics, group memberships under the two methods differ by 25 percent with 12 clusters, a number subjectively chosen based on screen plot and qualitative assessment.

Different configurations of PCA also yield noticeably different statistical efficiencies and associations. Using covariance instead of correlation to measure relationships achieves greatest efficiencies and generates the most meaningful output. Promax (oblique rotation) is used instead of Varimax (orthogonal rotation), recognizing that

economic activities may be correlated. As can be seen in the plots shown in Figure 1, correlation with Spearman and Pearson's generates murky principal components while covariance creates more logical constructs and greater explained variance.

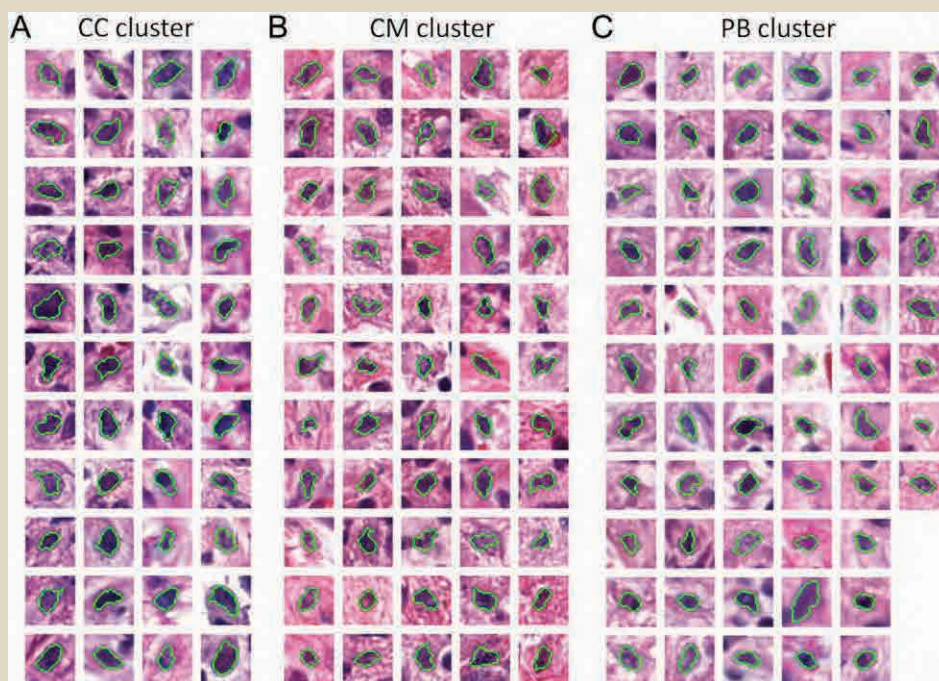
None of the techniques described in Figure 1 achieve particularly high levels of explained variance, in part reflecting the diverse nature of the global economy. However, the exercise illustrates how fundamental performance metrics and group formation can vary greatly even between commonly applied methods. And, of course, change the data set and the preferred methodology might change as well.

ANALYTICAL CONSIDERATIONS

There aren't any hard and fast rules about what data reduction method is optimal for a given situation. However, here are some general considerations to keep in mind:

1. Exploration and theory: Given the degree to which results can vary by technique, data reduction exercises are best treated as exploratory. Openness

to trying different approaches is recommended rather than relying on one method as the be-all and end-all. This doesn't negate the importance of having a theoretical foundation to shape discovery. Techniques such as cluster analysis and PCA will group and compress any data. Without a priori assumptions and hypotheses, analysis can turn into a wild goose chase.



Improving patient diagnosis and treatment

Cluster analysis can be used to classify diseases. This example groups glioblastoma using TCGA image data, coded professional assessment and patient records. Dr. Lee Cooper and coauthors write in the *Journal of the American Medical Informatics Association* that such frameworks have the potential to improve preventive strategies and treatments.

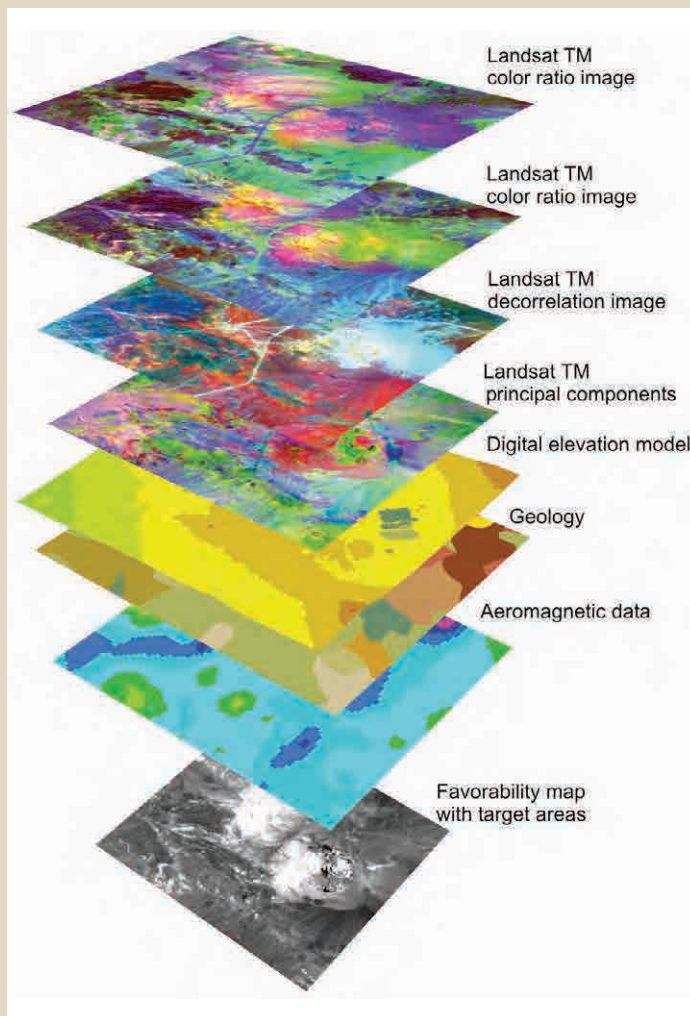
Source: [National Center for Biotechnology Information](#)

2. Sparsity and dimensionality:

Data with limited variation are likely to add little value to data reduction exercises and can be considered for exclusion. At the other end of the spectrum, too much dimensionality can complicate efforts. Discerning which attributes are most important up-front and preparing data accordingly play an important role in successful analysis.

3. Scale, outliers and borders:

Consistent scales are required so that variables with high magnitudes don't dominate. Outliers can also skew results. For example, because K-means partitions the data space into Voronoi cells, extreme values can result in odd groupings. At the same time, carte blanche removal of outliers isn't recommended given the insights that



Optimizing mineral exploration

Dimension reduction techniques enhance geological mapping exercises. In *Geosphere*, Norbert Ott, Tanja Kollersberger, and Andrés Tassara illustrate how principal component analysis with Landsat satellite data can improve mineral exploration efforts.

Source: [Geosphere](#)



EVERY BUSINESS... EVERY ORGANIZATION... AND EVERY ANALYTICS PROFESSIONAL...

Experiences the ups and downs, and the twists and turns of analytics. Making analytics work in real organizations can be a dynamic (dare we say wild?) ride for even the most seasoned practitioners. Analytics 2016 will help you conquer the challenge.

PRESENT A TALK OR POSTER AND SAVE!

SUBMISSION DEADLINES

ORAL PRESENTATIONS
January 4, 2016

POSTER PRESENTATIONS
January 8, 2016

SUBMIT OR REGISTER AT
meetings.informs.org/analytics2016

2016 CONFERENCE ON BUSINESS ANALYTICS AND OPERATIONS RESEARCH APRIL 10-12 IN ORLANDO

potentially lie in exceptional cases. Another challenge with K-means is the potential to incorrectly cut borders between clusters as it tends to create similarly sized partitions.

4. Order and iterations: Source data should be randomized if the algorithm used is influenced by the order in which observations are processed. And if the technique employed involves random start points rather than fixed or user defined, multiple iterations should be conducted to evaluate the extent to which results are consistent with consecutive runs.

5. Mechanics and validation: Failure to understand the mechanics of data reduction methods could result in a sub-optimal framework and misinterpretation. Being expert in all the different techniques is a tall task, but the wealth of information publically available makes learning as you go possible. With so many factors to consider, validation of results is also critical. Can results be substantiated statistically

and logically? What do results look like visually? Can results be replicated and effectively used for prediction?

We've only scratched the surface of data reduction, covering just a few of the many techniques at a very high level. However, we've shown that even some of the more popular methods can generate significantly different results. Data reduction clearly falls into the camp of "part art, part science." As such, it's probably fair to say that there's no "correct" approach. However, there are important considerations to be made when conducting analysis, perhaps the most important of which is the need to take an exploratory approach, open to the possibility that one size may not fit all. ■

Will Towler (wjtowler@hotmail.com) is an analytics and insights specialist. The views expressed in this article are those of the author and do not necessarily represent the views of an employer or business partners. He is a member of INFORMS.

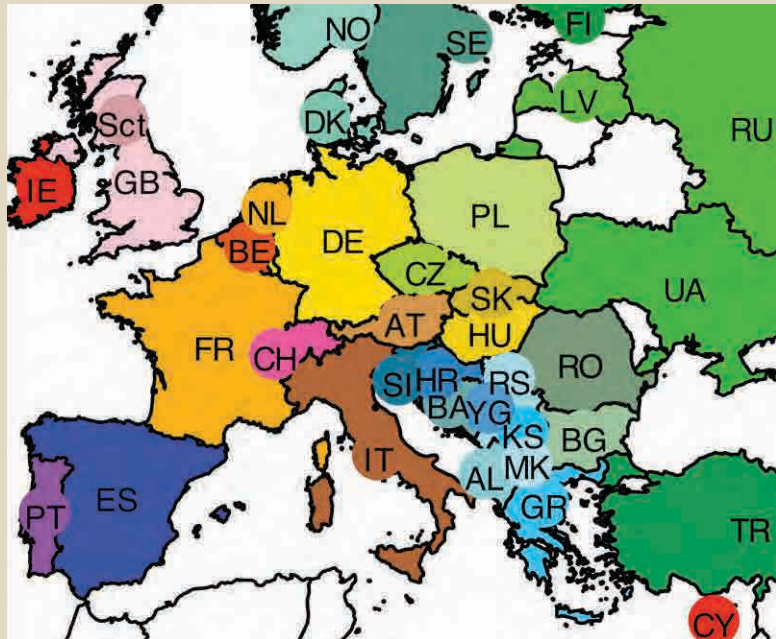
Request a no-obligation INFORMS Member Benefits Packet

For more information, visit: <http://www.informs.org/Membership>

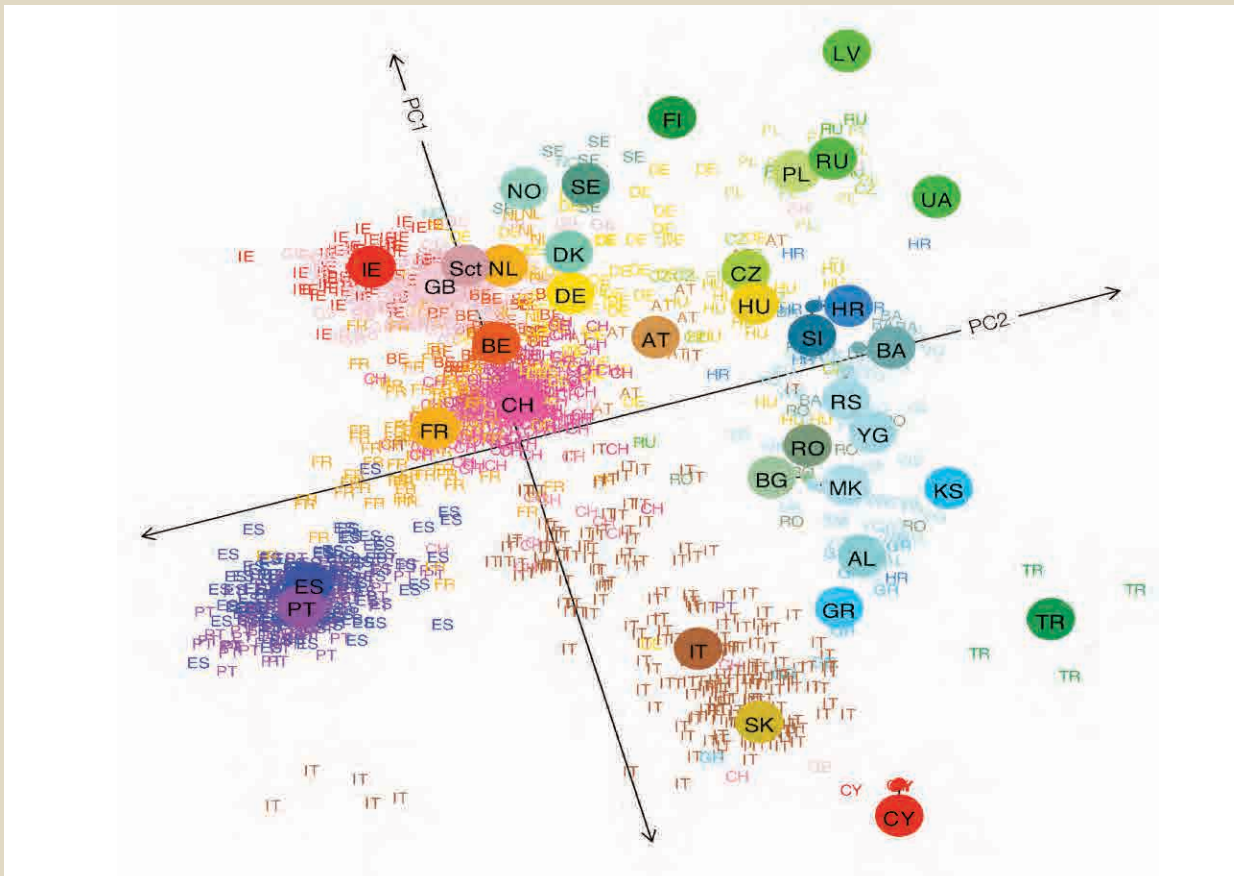


Decoding Genealogy

Dimension reduction can shed light on the genetic makeup of – and relationships between – human populations. In *Nature*, Dr. John Novembre and team highlight the close relationship between genetic and geographic distances using principal component analysis. They conclude that spurious associations can arise when conducting genetic mapping exercises if such considerations are not properly accounted for.



Source: *Nature*, Nov. 8, 2008; reprinted by permission from Macmillan Publishers.



'Ensemble of Ensemble'

Meet the new nominal in predictive modeling



BY ANINDYA SENGUPTA

Accurate prediction of the future is the most important motto of predictive modeling. With the adoption of predictive modeling and analytics across different industries, the focus on accuracy has increased massively. Predictive modelers across the globe have been in search of newer and more innovative techniques to increase their prediction accuracy. Against this backdrop, one method to consider for improving the accuracy of model predictions is the ensemble approach of modeling.

Ensemble models combine two or more models to enable a more robust

prediction, classification or variable selection. The most common form of ensemble is combining weak decision trees into one strong model. The most prevalent methods in this regard are bagging and boosting. These machine-learning techniques have better accuracy than the erstwhile traditional statistical techniques such as least square estimation and maximum likelihood. However, the traditional techniques are more robust. We should ideally have a blended approach wherein we integrate the best parts of the two models. The new buzzword phrase in the predictive analytics world is “ensemble of ensemble” models.

There are different ways of combining different models. The most simplistic way of combining models is to take a simple or weighted average of the predictions from different models. For business problems with binary target variables, such as predicting the likelihood of fraud, this approach sometimes gives better results in terms of predicting the event rate. The major limitation to this approach is that the overall accuracy of this combined model lies

between the accuracies of the individual models. Thus, this method is definitely not advisable for modeling continuous variables as overall accuracy matters most there. This method, however, can still be considered for modeling binary target variables.

BEST PRACTICES OF MODELING

Given the strength of the traditional models, one possible approach of combining the models can be to use

Upcoming Courses:

DATA EXPLORATION AND VISUALIZATION

NOV. 10–11, 2015

ESSENTIAL PRACTICE SKILLS FOR HIGH-IMPACT ANALYTICS PROJECTS

NOV. 18–19, 2015

Johns Hopkins University, Baltimore, MD



www.informs.org/continuinged

the traditional models for variable selection. Then, with the chosen set of variables, we should run the machine-learning method incorporating bagging and boosting. This will ensure that all best practices of modeling are maintained. One argument against using the machine learning techniques is that the variable selection process is not well defined. Two highly correlated variables can be selected in the model, and business may not have much insight separately from these two variables as both of them may be suggesting similar insight. One of the best ways to solve this kind of problem is to use traditional models for variable selection and use those variables for machine learning models. This can be one kind of ensemble.

One of the more prevalent forms of combining two different models has been to use the output of one model as input in the other model. This happens in two-stage models. The entire literature on censor modeling is based on these two-stage models. One can use the inverse mills ratio calculated from the output of the first-stage model as an input to the second-stage model. The method of using output from the model in stage one as an input in the model in stage two can also be treated as some kind of ensemble modeling.

One of the most efficient methods of creating ensemble of ensemble models is to use the predictions of different models to develop a separate model on the target variable using the predictions as the predictors. In this way, we will get the final predictions as a function of the predictors. The model will determine the functional form. In this method, we are basically optimizing the predictions using the individual predictors. This method gives better results in terms of the overall accuracy of the model. The combined model generally has higher accuracy than the individual models.

MORE ACCURATE RESULTS

Whatever way we use to combine different models, it has been generally observed that “ensemble of ensemble” models give more accurate results than individual models. This has been the new normal in the predictive modeling world.

Going forward, ample research is needed on exploring more innovative ways of creating ensembles of traditional and the machine-learning models. Given the focus on prediction accuracy, researchers need to focus on more scientific ways of combining models in order to ensure maximum accuracy. ■

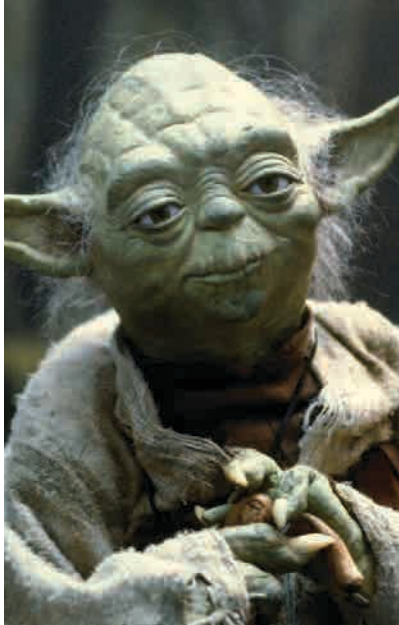
Anindya Sengupta is an associate director at Fractal Analytics (www.FractalAnalytics.com).



CERTIFIED ANALYTICS PROFESSIONAL

Analyze What CAP® Can Do For You

www.certifiedanalytics.org



Jedi Master Yoda: Let the data be with you.

Star Wars analytics II: Predicting the new movie's sales

I'm doing something that I have never done before, which is to write about the same topic twice in a row. The new Star Wars movie is that big of a deal. The previous installment considered the [Battle of Hoth](#).

I remember the first movie I ever went to: "Star Wars: A New Hope." We knew it back then as just "Star Wars." Last month, I focused backward, on events in "Empire Strikes Back" (another in the Star War series). This month, I focus forward, thinking about sales of the next Star Wars movie.



**BY HARRISON
SCHRAMM, CAP**

Analytically, the first thing to try is prediction based on regression. The data we have at this time is six Star Wars movies with which to compare [1]. Given the data, we are tempted to simply perform a linear regression to predict the box office of the new Star Wars movie as shown in Figure 1.

The F-statistic associated with this regression has a significance $p = .102$, which is probably OK for this application. So, we're like, done, right?

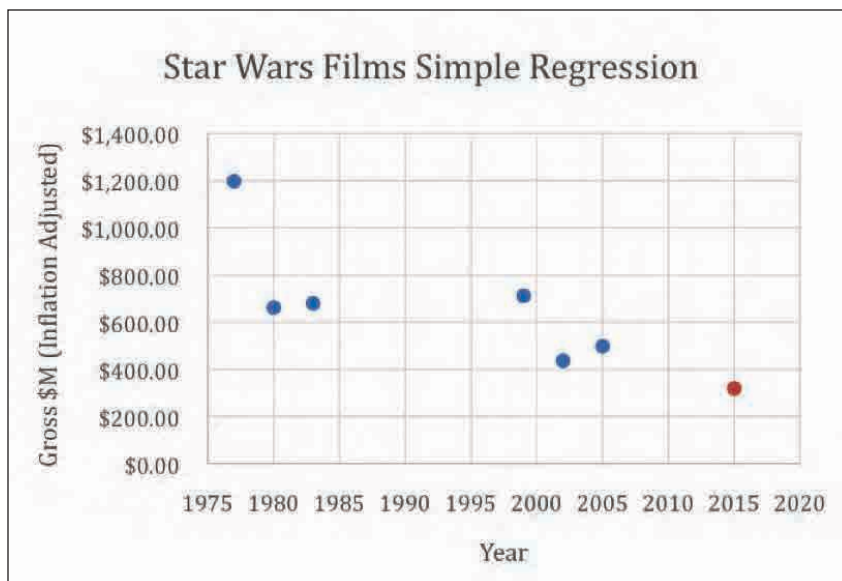
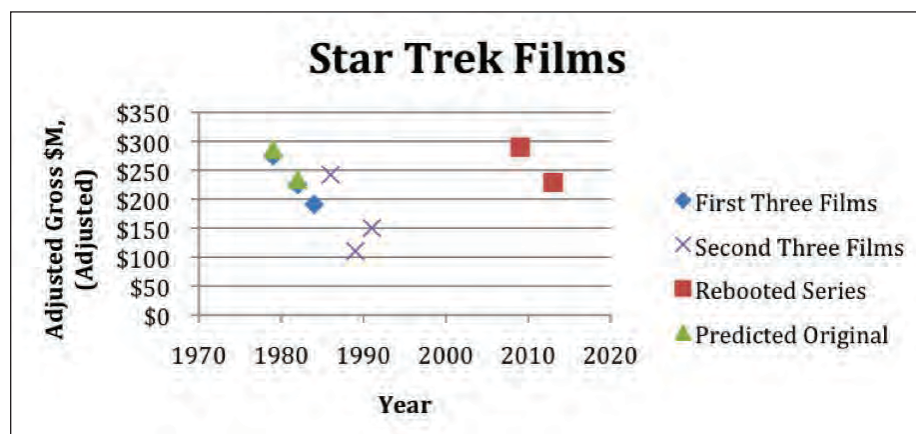


Figure 1: Simple regression of the Star Wars movies. Episodes 1-6 are represented by blue dots, and Episode 7 (predicted - trend) is represented by a red dot.

Figure 2: Comparison of Star Trek films, by series. The “rebooted” series, consisting of “Star Trek” and “Star Trek Into Darkness,” made approximately 4 percent more than the original series, in inflation-adjusted revenue.



Wrong!

With apologies to famed Star Trek character Mr. Spock, I find the application of analytic methods without consideration of the underlying problem highly illogical.

Market forces can be complex, and I do not believe without proof that the new Star Wars movie will perform simply as an extension (on trend) from the previous six. It is also worth noting that the original Star Wars movie was the No. 2 grossing

film of all time (behind “Gone with the Wind”), and that all six films are in the top 50 grossing of all time.

I recommend at this point a comparison, and this might be controversial and surprising. There is a science fiction franchise with astonishingly similar history to include multiple restarts: “Star Trek” (see Figure 2).

I’m going to argue that the new Star Wars movie has more in common with the rebooted Star Trek than laser-pistols.

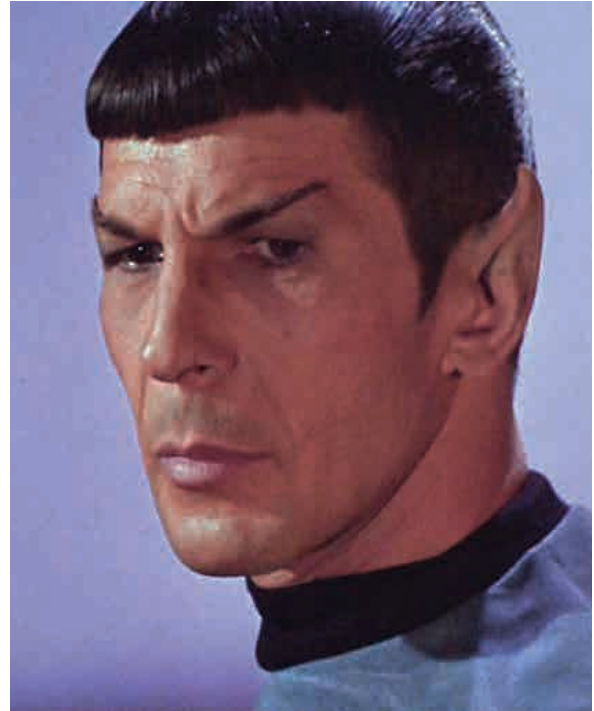
First, they both have similar spacing between films. Second, both the original and rebooted series will feature actors from the first: Leonard Nimoy (who sadly passed away this year) in Star Trek and at least Harrison Ford in Star Wars. Finally, the films have multi-generational appeal; first with kids, who are experiencing the series for the first time and will drag their parents, and parents who grew up with the original films and will drag their children. Finally, both the Star Wars and Star Trek reboots are directed by J.J. Abrams.

Channeling my inner Yoda of Star Wars fame, consider this: Data is all around you. Operations researchers feed on it. Statisticians breathe it. Strengthens them it does.

Using this comparison, we would inflate the (original) Star Wars by 4 percent and then compensate for inflation, and come up with \$1.55 billion. This puts our estimate of the new Star Wars movie squarely between the original Star Wars and the No. 1 grossing movie of all time, “Gone With The Wind.” Did I really just predict that “The Force Awakens” will be the new No. 2 movie of all time, outgrossing the original after compensating for inflation?

Yes, I did.

Open issue: the sophomore slump. It is noteworthy that for all four sets of



Would Spock find analytics “fascinating”?

movies considered (Star Wars IV-VI, Star Wars I-III, Star Trek I-III, Star Trek IV-VI) that the middle movie of the trilogy performed worse than the first and last. I don’t have an explanation for this. Also, it seems (and I have predicted) that the middle trilogy performs worse than the first and last. ■

Harrison Schramm (harrison.schramm@gmail.com), CAP, is an operations research professional in the Washington, D.C., area. He is a member of INFORMS and a Certified Analytics Professional.

REFERENCES

1. All data from www.boxofficemojo.com, accessed early October 2015.

ACCESS YOUR COPY!
2015 TutORials available ONLINE
for INFORMS Members!

TUTORIALS^{IN}

OPERATIONS RESEARCH

The Operations Research Revolution

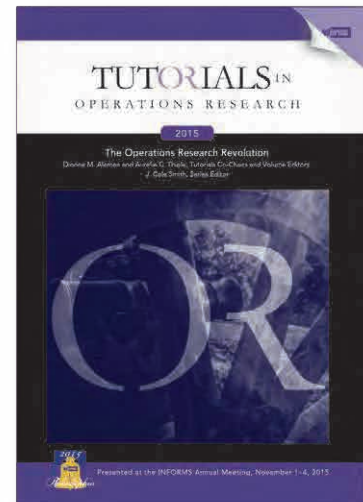
Dionne M. Aleman and Aurélie C. Thiele, Tutorials Co-Chairs
and Volume Editors

J. Cole Smith, Series Editor

INFORMS 2015 edition of the *TutORials in Operations Research* series
will be available online to registrants of the 2015 INFORMS Annual
Meeting on November 1, 2015.

Access the 2015 *TutORials* at:

<http://pubsonline.informs.org/series/educ>



informs

SAVE THE DATE

JUNE 12-15, 2016

HILTON WAIKOLOA VILLAGE

**SAY ALOHA TO AN UNFORGETTABLE
INFORMS INTERNATIONAL MEETING IN HAWAII**

TOP FOUR REASONS

**YOU SHOULD ATTEND THE
INFORMS INTERNATIONAL CONFERENCE:**

- 1) Network, Connect, and Reconnect with Colleagues
- 2) Learn from One Another and Develop Professionally
- 3) Refresh and Recharge
- 4) Location, Location, Location!



<http://meetings2.informs.org/wordpress/2016international>

Racecar design

Table 1: Engines and tires and brakes, oh my!

Sample car configuration	Engine	Tires	Trans mission	Brakes	Race finish speed (minutes)
1	Awes Engine	Aero Tires	Accelo Trans	Control Brakes	12
2	Charm Engine	Boldo Tires	Accelo Trans	Control Brakes	17
3	Charm Engine	Aero Tires	Beast Trans	Bolt Brakes	14
4	Baller Engine	Boldo Tires	Accelo Trans	Control Brakes	13
5	Awes Engine	Boldo Tires	Beast Trans	Aero Brakes	15
6	Awes Engine	Clingo Tires	Beast Trans	Control Brakes	11
7	Baller Engine	Clingo Tires	Cert Trans	Aero Brakes	16
8	Baller Engine	Boldo Tires	Accelo Trans	Bolt Brakes	19
9	Charm Engine	Aero Tires	Cert Trans	Aero Brakes	18
10	Charm Engine	Clingo Tires	Cert Trans	Bolt Brakes	20

As the lead design engineer on a car racing team, you've been hired to build a car that can complete a racecourse in the fastest time. To get a feel for how each component performs, you've built several sample car configurations with varying engines, tires, transmissions, and brakes and then tested them on the race-track. Each combination of parts has yielded different speed results, as shown in Table 1.

There are three types of each component that you can choose from. You must pick exactly one component from each category: engine, tires, transmission and brakes.

BY JOHN TOCZEK

John Toczek is the senior director of Decision Support and Analytics for ARAMARK Corporation in the Global Operational Excellence group. He earned a bachelor of science degree in chemical engineering at Drexel University (1996) and a master's degree in operations research from Virginia Commonwealth University (2005). He is a member of INFORMS.

Question: What combination of engine, tires, transmission and brakes will give you the fastest car?

Send your answer to puzzlor@gmail.com by Jan. 15, 2016. The winner, chosen randomly from correct answers, will receive a \$25 Amazon Gift Card. Past questions and answers can be found at puzzlor.com. |

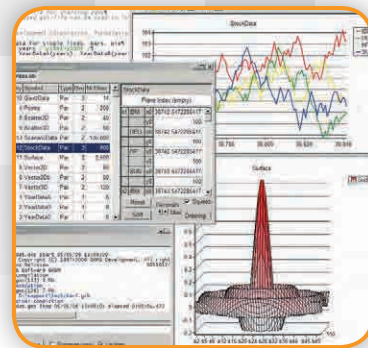
GENERAL ALGEBRAIC MODELING SYSTEM

High-Level Modeling

The General Algebraic Modeling System (GAMS) is a high-level modeling system for mathematical programming problems. GAMS is tailored for complex, large-scale modeling applications, and allows you to build large maintainable models that can be adapted quickly to new situations. Models are fully portable from one computer platform to another.

State-of-the-Art Solvers

GAMS incorporates all major commercial and academic state-of-the-art solution technologies for a broad range of problem types.



GAMS Integrated Developer Environment for editing, debugging, solving models, and viewing data.

Long-term Energy Scenarios in the US with TIMES FACETS

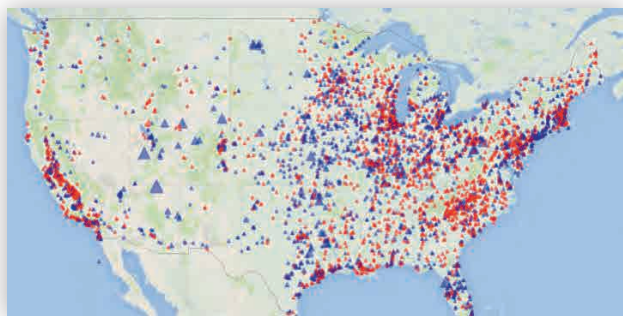
Making decisions about energy policy requires assessing impacts on many intersecting goals – including energy security, climate change mitigation, air quality, economic development, and energy access – in the context of significant uncertainties about future fuel markets, technology development, and policy choices. The TIMES model generator, written in GAMS and stewarded by the International Energy Agency's Energy Technology System Analysis Program (ETSAP), a consortium of 20 research institutions worldwide, provides building blocks to depict an entire energy system from "wells-to-wheels" as a network of technology pathways, fully capturing supply-demand interplay.



Explore the solution space



Reach policy makers



Model at many scales

In the US, the TIMES FACETS model represents individual power plants, along with their fuel supplies, emissions, and retrofit options. FACETS has been used to examine mercury emissions control, the economic impacts of shale gas, and most recently the EPA's Clean Power Plan. It provides state-level results within regional and national contexts. Use of a cloud-based GAMS platform allows hundreds of model runs to be submitted simultaneously, enabling full exploration of the solution space. The VedaViz data visualization system and policymaker portals help extract insights from the large volume of results data and engage a wide range of stakeholders in the analysis process.

For more information see www.iea-etsap.org and www.facets-model.com.