Robert L. Mitchell

Reality Check

More posts | Read bio

August 10, 2011 - 11:11 A.M.

How Macys.com visualizes the future

Add a comment

What's the likelihood that a loan applicant will default? What mix of sizes, styles and colors will sell best in a given store? Predictive analytics can provide forecasts that work like magic, but the results are only as good as the models that underlie them. Those models are developed by people like Kerem Tomak and his team at Macys.com, the online business unit of the Macy's department store chain.

Predictive analytics didn't exist a year ago at Macys.com but is quickly gaining momentum, says Tomak, vice president of marketing analytics.

His predictive analytics process, built using tools from SAS Institute, goes through four stages: discovery, data visualization, modeling and finally validation and testing. The data discovery and visualization steps guide the modeling approach, and being able to visualize the problem is critical, he says.

"The choice of the right predictive model is what separates a successful modeler from a not-so-successful one," he says. To build a successful model, analysts must understand the data and how the data or the phenomenon they are trying to capture behaves. **"When building a model there are lots of choices to make, ways certain functions or behaviors can be written in mathematical terms,"** he says. To assist in this process, Tomak says he relies on tools that help him visualize the problem.

"It's easier to understand a phenomenon visually."

Tomak's tools include SAS data mining and modeling tools that sit on top of Hadoop, the open source distributed computing environment in which most of Macys.com's analytics data resides. Tableau Software and SQL handle the visualization piece for large data sets, which is most of the data. Analysts use SQL and Excel for smaller data sets.

Macys.com uses visualization for business and customer insights as well as a way to set the stage for predictive analytics modeling. For example, Macys.com uses visualization to view trends in customer traffic coming into its Web site, such as how product families are performing and how sales of products track around different marketing events.

Tomak's **ultimate goal is to have one centralized analytics portal** through which all parts of the business can share data and insights, built on top of one data warehouse that connects to all of the data sources in the business. "All of the data used by the analytics team is in Hadoop but it's in different formats in different locations," he says.

But if having one central data mart -- one source of truth -- for all of the business' entities is a laudable goal, it isn't always practical because technology continues to evolve and new tools are always being brought in to keep up. "I've been in this business since 1988 and it never seems to happen," Tomak says.

"There are always multiple tools and multiple databases. That never goes away."