Data Science 2.0 – Guided and In-line Analytics

Michael O’Connell, PhD
Industry Analytics
TIBCO

In retail banking and capital markets, IT initiatives and business projects are justified by improving productivity, reducing risk and/or growing revenues. The various functional areas of these organizations rely on big data analytics – trading and managing portfolios, creating value across customer relationships, detecting and preventing fraud and managing risk across the enterprise.

Data Science is the practice of deriving insights from data to solve business problems. The current market wave is tied to the new world of big data, and Data Scientist is now touted as the sexiest job of the 21st century in the Harvard Business Review [1]. McKinsey has noted a 50-60% supply-demand gap for data scientists, with a shortage of more than 150,000 data scientists and 1.5 million managers with big data analytics understanding over the next 5 years [2].
The value of data science is undeniable, with core applications driving a wide array of business and customer intelligence programs including:

- acquiring and growing customers (cross-sell / up-sell)
- attrition modeling (with intervention!)
- trigger-based marketing (including mobile and location based offers)
- sales lift analysis (test and learn)
- dynamic segmentation (treating different segments differently!)
- pricing analytics (including loyalty and lifetime value)
- sentiment and attribution analysis (staff performance and advertising/campaign effectiveness)
- fraud and risk analytics (operational execution and enterprise analysis)

In addition, specialized data science programs drive industry vertical solutions, with industry-specific data sources that are expanding and evolving e.g. to incorporate social networks and real-time data. In the Financial Services market, this plays out across all of the core functional areas including:

- capital markets (trading, portfolio)
- retail (market and customer insight)
- risk management (enterprise, credit and counterparty, market)
- fraud and compliance (AML, credit card, trading)

With the current excitement around big data and data science there are some solutions getting attention that are dangerous e.g. promoting isolated, end-user, database analyses with all the attendant problems such as chasing noise, throwing out data that don’t match pre-conceptions, confusing leading and lagging indicators, interpreting correlations as causation e.g. as described by Silver [3]. Some new companies are even making bold claims that they invented visual analytics, that data scientists will be dead in 18 months; or that we don’t need data scientists, just easier access to big data. This type of thinking introduces substantive risk – not only of heading in entirely the wrong direction, but with significant negative ramifications, like bringing down businesses, operations or financial systems! The end-user community can be readily enabled with self-service analytics (as outlined below), but there needs to be inbuilt guidance, and a framework of end-user data discovery, collaboration and enterprise readiness that promotes rigorous and real analysis on the business!
I outlined the basic Data Science 1.0 process in a 2012 Forbes article, “What is a Data Scientist” [4], and in notes to the European Banking Forum [5] and WSTA [6] communities. With recent technology innovations [7, 8], we have now jumped beyond this to what I believe to be a new Data Science 2.0 state. The expanded workflow in this new state includes:

- identifying the high-value business problems and developing value theses with demonstrable ROI
- assembling the appropriate data mashups to address the problems
- ordering the data aggregations and filters – in-database and in-memory
- exploring the data (EDA) – visually and interactively
- constructing and validating the features that inform the problem – leading and lagging indicators
- deploying the feature sets and exploratory data analyses as self-service, guided, collaborative analyses across all relevant functional areas in the enterprise – with elastic architectures to efficiently meet demand
- building and evaluating models that describe and/or predict the measured response
- deploying the champion model in the real-time event system driving the business solution across the customer and market space
- building and evaluating new features, dashboards and challenger models for evolution of the guided analyses and in-line event analysis systems
This workflow is illustrated in Figure 1. for a typical business insight use case in financial services.

Figure 1. Data Science 2.0 workflow for a typical business insight use case in financial services.

In the Financial Services market, our businesses are evolving rapidly and we are working hard to be nimble and intelligent with our information solutions. This combination of interactive, visual, descriptive and predictive analytics; with self-service guided and collaborative workflows for the masses, and in-line deployment in real-time event systems, is the future. This is Data Science 2.0.
References:


Michael O’Connell, PhD, Industry Analytics, TIBCO Spotfire (http://about.me.com/moconnell). Please visit http://spotfire.tibco.com for more information.