Open Source Software Fuels a Revolution in Data Science

March 16, 2015 by Daniel Gutierrez Leave a Comment

In this special guest feature, Neera Talbert of Revolution Analytics *discusses the role of open source software in making data science the rising field it is today. Neera Talbert is Vice President, Professional Services at Revolution Analytics, a statistical software company focused on developing open source and "open-core" versions of the free and open source software R for enterprise, academic and analytics customers. Neera is a seasoned consulting executive with over 25 years of experience.*

It's hard to downplay the influence of open source software on the spectacular rise of data science. From my perspective as a technology consultant, open source isn't just an interesting aspect of the data science revolution; it's absolutely critical.

R, a programming language originated in 1993 by two academics in New Zealand, is a great example of the power of the open source community on the global economy. Conceived specifically for statistical data analysis, R has played a major role in elevating the practice of analytics to its present state, and it seems likely to continue as a propulsive force in this rapidly growing field.

The rise of data science and the role of R in fueling that ascent make it imperative for schools and universities to revisit their curricula in at least three areas of study: computer science, statistics and business.

Why those three areas? For the answer, let's look at the role of the modern data scientist. Unlike a pure statistician, a data scientist is also expected to write code and understand business. Data science is a multi-disciplinary practice requiring a broad range of knowledge and insight. It's not unusual for a data scientist to explore a fresh set of data in the morning, create a model before lunch, run a series of analytics in the afternoon and brief a team of digital marketers before heading home at night.

In addition to possessing a wide range of practical knowledge, a data scientist must also be agile and flexible. Today's swiftly changing markets require lightning fast reflexes – companies must be capable of assessing new data and responding in the space of a heartbeat to unexpected shifts in commerce, across all industry verticals and economic sectors.

The speed of modern business plays to the strengths of data science and open source programming. In the past, business moved relatively slowly and large-scale market trends were fairly predictable. As a result, most companies were quite comfortable relying on proprietary (closed source) software to analyze data. The downside of proprietary software, however, is that it cannot be quickly modified or updated to handle unexpected circumstances or disruptions of existing business models. Until recently, it was common practice for traditional vendors to release updated versions of critical proprietary software quarterly or annually.

Open source software can be modified or rewritten in days or hours, making it an ideal choice for real-time analytics. The global R community also generates tools and statistical packages that can be downloaded at no cost, giving data scientists a virtually inexhaustible supply of fresh programming resources.

Moreover, the open source movement is democratizing data science. In the past, you needed special training on a proprietary system and years of experience to become a valuable member of a business or research team. Thanks to a wider choice of open source tools, more people can begin contributing valuable insight and analysis from the start.

I encourage any student who is interested in computer science, statistics or business to learn as much about R as possible. I

also urge schools and universities to offer classes and instruction in open source programming. The multi-disciplinary nature of the modern economy requires all of us to look beyond traditional disciplines and develop new skills. I know there's a lot of talk about the need for specialization, but data science welcomes people who are genuinely interested in the world around them.

I've had a wonderful career in the technology industry, and from my point of view, our best days are still ahead of us. The combination of data science and open source programming opens up a new universe of opportunities at many levels and in many places. Let's grab those opportunities and run with them.