Data Mining Blog by Salford Systems

SALFORD SYSTEMS
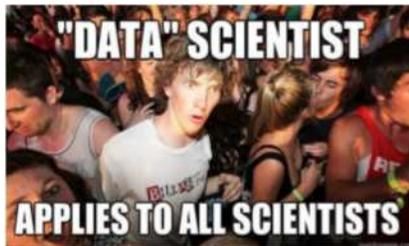
| RSS Feed

Musings on Becoming a Data Scientist [guest post]

Posted by Heather Hinman on Fri, May 09, 2014 @ 09:25 AM

Guest Post by Scott Terry, Rapid Progress Marketing and Modeling, LLC

If you're new to this great industry, when you first heard the term "Data Scientist", maybe you were seduced by the mystique, the buzz, and you dreamed of all the things you could achieve with this powerful new science (including perhaps, world domination.) On the other hand, if you've been around a while, you probably wondered what the heck was going on because, as the great Philosopher Yogi Berra once said, "This is déjà vu all over again."

Since I'm in the latter category, imagine waking up a few years ago to hear that what you've been doing for decades is new. Really? Is it? Then what exactly is a Data Scientist and how do you get to be one? Since my curiosity was piqued, I initiated a quest to see what it was about.

Data Scientists do Data Science (and my dog is happy to see me)

My first effort to discover what a Data Scientist was about came through a consultation with my good friend Google. No, I don't mean the company. Though I do work late nights, I'm not young enough, I'll never confess to eating Hot Pockets, and well, I'll never be on their radar screen. (Yes, I know. They have tons of data about me, and you, and practically every other person on the planet.) You know what I mean … I mean the search engine.

After some time searching on Data Science, to my dismay, I found nothing new. No fresh meat. Zero. Zip. Nada! Though people were chattering about it in a bazillion places, apart from a Data Scientist doing Data Science, it seemed that nobody really had a real, consistent handle on what that meant. It reminded me of my 65-pound lap dog, Zeus. When I step out to the mailbox for ten seconds and return, he's excited to see me. And then, his excitement gets me excited though neither of us are quite sure why. Similarly, I learned one thing was for sure … everybody was getting excited about Data Science and it was spreading!

More on Moore (not less)

At this point, I have to confess, I started to warm up to the title a little. After all, being a "Scientist" is sexier sounding than "Modeler" or "Analyst." And while all of this was sinking in, I was seeing another term pop up with increasing frequency. That term was "Big Data." And it was usually being paired with "Data Scientist."

Since I'm not as smart as Archimedes, I can't claim it was a "Eureka!" moment. Rather, it was more of an "Aha!" moment. The ~~light bulb~~ *curly fluorescent tube* went off. "Hey! Someone thinks Big Data is new!" For sure, if you're new to the industry, indeed, it would seem new. But from where I look at the industry, it's not new. It's really just the next point on a continuum. Let me explain.

You've heard of Moore's Law? It's the one that says that the number of transistors on integrated circuits doubles every two years. So far, it's been a pretty accurate phenomenon but what it means for us is that the geometric progression of computing power has had a commensurate impact on Big Data. It allows us to work with more data but that same power also has ignited the generation of more data. *Tech is both a cause and the effect of Big Data!*

Forty years ago, we lived on the cusp of personal computing. Just prior to that time, believe it or not, "Big Data" was created when the rubber band broke on your stack of Hollerith punch cards and they fell in random disarray on the floor. Then, with the invention of the integrated circuit, the revolution of ever-scaling microprocessing began.

Thirty years ago, Big Data became gigabytes, twenty years ago it was terabytes, and today it's petabytes. So if you think what we're working with today is "Big" Data, wait another ten years and look back to this time. When you do, you'll have your own "Aha!" Congratulations … you've joined the club.

You see, neither Data Science nor Big Data are really something new. They're ongoing strands of continuous development. Plus, you and I are always riding the peak of the "Big Wave." So hold on because the wave is only going to get even bigger, thanks to technology.

What Does it Take To Become a Data Scientist? (and play it again Sam)

That's THE important question, isn't it? And the answer is … what was old has become new again.

With no allusions to La Cosa Nostra, when I received my degree in "our thing," it was called "Quantitative Arts and Sciences." It was descriptive then and it's still descriptive today. The coursework laid a foundation of technology (most notably systems and programming) and it integrated applications throughout. Those applications were Statistics, Marketing Research, Operations Research, and Artificial Intelligence. And today, guess what? **To become a Data Scientist, you still need to achieve good measures of fluency in the big three:**

1. **Technology,**
2. **Statistics, and**
3. **Artificial Intelligence.**

But as time as taught, it takes even more than that.

You see, it's more than Science. As the title of my degree implied (though I really didn't understand it at the time), perhaps the largest portion of what it takes to become a Data Scientist involves learning the "Art" of "our thing." It's not about total bytes, it's about the total value you produce from the data. And to produce that value my friend … that's the art.

Think of a concert pianist. In order to be the best at what they do, they must have knowledge of music. That's analogous to us learning the "big three". But that's not enough. A pianist's process input is the written piece of music. That's exactly like our data. We polish and prepare it as input to our instruments to play them. And the instruments? Instead of pianos, our instruments are software. But concert pianists don't play just any instrument, they want play the best instruments … cutting edge tools like Salford Predictive

Modeler.

And even with all of these elements in place, the concert pianist must continue to learn and practice to learn their art and get better at it. Once started, it never stops. You keep progressing down the continuum. Now, let me ask you … have you already heard that tune here? By now, does it have a familiar sound?

So, if you ask me about what it takes to become a Data Scientist, I will tell you it's a process that blends acquired knowledge and skills to deliver value. It's an art that I have been practicing for decades and will never stop. And if you ask me this same question again in 20 years, I'll likely tell you that I'm still becoming a Data Scientist. By then, you will, too.

Scott Terry is President of Rapid Progress Marketing and Modeling, LLC (RPM2) of St Petersburg, FL, a former Fortune 1000 Executive, entrepreneur, and Data Science veteran of 30 years. Recently named one of the "Top 100 Most Promising Big Data Companies" by CIO Review, RPM2 was founded in 2008 and specializes in providing businesses with comprehensive Predictive Analytics services, Consulting, and Data Science Education.

Tags: **big data**, **data science**

Post Comment

Name

*

Email

*

Website (optional)

Comment

*

Allowed tags: <a> link, <b> bold, <i> italics

☐ Receive email when someone replies.

☐ Subscribe to this blog by email.

Post Comment