

KDnuggets

[Subscribe to KDnuggets News](#)[Contact](#)

- [SOFTWARE](#)
- [News/Blog](#)
- [Top stories](#)
- [Opinions](#)
- [Tutorials](#)
- [JOBS](#)
- [Companies](#)
- [Courses](#)
- [Datasets](#)
- [EDUCATION](#)
- [Certificates](#)
- [Meetings](#)
- [Webinars](#)

[KDnuggets Home](#) » [News](#) » [2017](#) » [Feb](#) » [Opinions, Interviews](#) » KDnuggets
Top Blogger: An Interview with Brandon Rohrer, Top Data Scientist ([17:n06](#))

KDnuggets Top Blogger: An Interview with Brandon Rohrer, Top Data Scientist

[◀ Previous post](#)[Next post ▶](#)

http likes 64

Tags: [Brandon Rohrer](#), [Data Science](#), [Top bloggers](#)

Read an interview with Top KDnuggets Blogger Brandon Rohrer, and get his thoughts on data science, newcomers to the field, and his ambitious pet project.

By [Matthew Mayo](#), KDnuggets.

As the next entry in our series of interviews with KDnuggets Top Bloggers, we speak with well-recognized data scientist and prolific blogger [Brandon Rohrer](#).

Brandon does a great job of introducing himself to readers below, so I will skip that here. I will point out that Brandon can be found on [LinkedIn](#) and [Twitter](#), and he maintains his blog [here](#).

So read on to get insight into Brandon's thoughts on data science, what newcomers to the field need to know, and an impressive (and ambitious) pet project.



Matthew Mayo: Hi Brandon! First off, thanks for taking time to speak with our readers. Why don't you introduce yourself and give us a little bit about your professional background.

Brandon Rohrer: Hi, my name is Brandon Rohrer, and I work at Facebook as a data scientist. My path has been a circuitous one. I had a childhood fascination with robots that was born from watching Luke Skywalker's prosthetic hand in the Empire Strikes Back. I played with computers and automobiles in high school, got an undergraduate degree in mechanical engineering, and studied robots for my masters and PhD at MIT. That was followed by a decade-long research career at Sandia National Laboratories and then a transition to industry through agriculture and tech.

How did you get into data science? Was it deliberate, or was it the happy accident that seems to be a recurring story for a lot of other successful folks in the field?

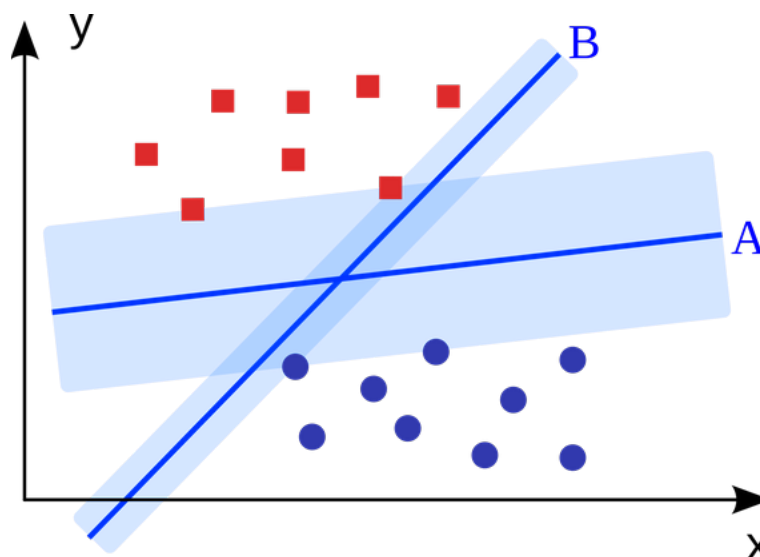
The metamorphosis from robotics researcher to data scientist was organic. Robotics requires the integration of different types of information. Signals can be noisy, sensors can be broken and the data can be extremely difficult to interpret. Robotics involves computer vision, signal processing, navigation, object recognition, decision making and all incarnations of machine learning. When you strip away the robots, the underlying problems look a lot like business and product problems that industry data scientists grapple with everywhere.

A friend helped me discover the field of data science. He works as a professor, and one day when I was talking with him he pointed me to a job posting at an agricultural company. I realized that I was a good match for the skills required, and was surprised to find the job title was data scientist. I got the job and decided that I absolutely loved it. Coaxing insights and decisions out of big collections of numbers and values continues to hold a fascination for me bordering on addiction.

Along with your day job, you are also a prolific data science blogger. Where do you get the motivation to blog as much (and as informatively) as you do, and how do you decide on topics to write about?

Of all the classes that I took, the most useful professionally were those for my creative writing minor. While writing short stories I learned some fundamental ideas like writing for your reader and editing ruthlessly. Those more than any other skill have been useful in my career. I've found that I really deeply enjoy, if not the process of writing itself, then creating a written work at reaches someone or in some way sparks their understanding. This has motivated me to devote some of my free time to capturing answers to questions and sharing them around. I've also found that writing is a great way to talk to more people. There are some questions that I get asked often. It's very effective to set aside some quiet time, reflect and write a thoughtful answer. Then I can share that with people who ask me the same question in the future. I also use tutorial writing to help me understand machine learning concepts. Some of my most popular blogs, like [How Convolutional Neural Networks Work](#) and [How Bayes Law Works](#) came about because of interview questions that I was asked and couldn't answer. I decided I wanted to learn those topics well, so I wrote the tutorials I would need to teach them to myself. The result I've captured and shared with others who might be in the same boat.

I have been somewhat surprised at how little truly introductory information exists on machine learning and data science topics. It seems that there is a natural tendency when we are familiar with the topic to forget what it's like not understand it. When I'm writing a tutorial I imagine that I'm explaining it to one of my children. That helps me to avoid using words and concepts that are unnecessarily complex, and when I do need to use them I take the time to define them in language that a 12-year-old would understand.



Just for fun: What's your favorite machine learning algorithm, and why? :)

I think of learning algorithms as tools in my toolbox. I don't have strong emotional attachments to anyone of my screwdrivers, but when I'm doing a job and I find just the right one for the job it makes me happy. I don't have favorite algorithms, but I do have favorite problems. We humans do things all the time that machines find extremely challenging. I interact with my world, receiving a massive amount of multimodal sensory information, but no labels or direct symbolic inputs of any sort. It's up to me to create my own set of symbols that meets my needs and to make decisions in a time constrained environment with partial and sometimes flawed information. And to do this well, I have to create complex models of the world, other people and myself, and I have to update them in a reasonable way on an ongoing basis. This set of problems, sometimes labeled the big AI problem or artificial general intelligence, capture my imagination like no others. I love that there are no tools that seem capable of solving this problem at the moment. That is a powerful motivator for me to join forces with the other people who have been enchanted by this challenge, to work on the edges of this problem and try to make a bit of headway.

There are lots of smaller problems that I love as well. Anything that requires modeling of a complex system, such as how corn grows or how the brain works or how large network of people interact and make decisions. Existing naïve methods don't seem to do very well on complex problems like these. Folk wisdom and domain knowledge is a useful starting point, but incomplete. I deeply enjoy the process of codifying domain knowledge as much as possible and building it into a learning approach that is flexible enough to unlearn the few pieces of domain knowledge that happened to be counterfactual.

In your opinion, what is the number one thing that new data scientists overlook or underestimate when entering the profession?

There are two things that are easy to overlook when entering the data science profession. The first is clichéd: communication. A great data scientist is a bridge between technical people (software engineers, statisticians, machine learning experts) and non-technical people (program managers, business development experts, corporate leaders, and the public). The ability to tell a story or convey information in an understandable way is not just a plus, it's essential.

The second thing that's really easy to overlook is more subtle. It is that it is important to understand the stories behind your data. How was each data point measured? Where they all measured by the same person? Are you certain that they mean exactly what you think they mean? How are they defined? What filtering or preprocessing has been done? Are the details of the methodology the same in every case? Are there systematic differences in how they're sampled? And when they were measured? What assumptions were made? Every single number has a story behind it. Of course you can't learn all their stories. But every minute of time that you invest learning a few of their stories will be paid back many times over in dead ends and misinterpretations avoided, and shortcuts taken to the answers that you need.



Can you tell us a little bit about [Becca](#)?

Becca is the current state of my efforts to solve the artificial general intelligence problem. It's researchy and still only handles toy environments, but I'm proud to share it around and let anyone who wants to kick the tires. It has been sitting on the shelf for the last little while, but I hope to return to it again at some point in the not-too-distant future.

More relevant to this question, Becca has been my vehicle for learning Java and Python. It's been my tool for learning classes and object oriented programming. It is been my path to understanding numerical computation and to implementing my very own variations of deep neural networks and reinforcement learning. The piece of advice I give to anyone looking to become a data scientist or to become a better data scientist is to build things. Becca is what I'm building.

On behalf of our readers, Brandon, I would like to thank you for taking the time to answer our questions today, as well as express our continued support and gratitude for the great content you consistently share with our readers, as well as on your own blog and beyond.

Brandon Rohrer recent KDnuggets posts include:

- [How Bayesian Inference Works](#) - 15 Nov 2016
Bayesian inference isn't magic or mystical; the concepts behind it are completely accessible. In brief, Bayesian inference lets you draw stronger conclusions from your data by folding in what you already know about the answer. Read an in-depth overview here.
- [How Convolutional Neural Networks Work](#) - 31 Aug 2016
Get an overview of what is going on inside convolutional neural networks, and what it is that makes them so effective.
- [Data Science for Beginners 2: Is your data ready?](#) - 28 Jul

2016

This second video and write-up in the Data Science for Beginners series discusses what is required of your data before it can be useful.

[◀ Previous post](#)
[Next post ▶](#)

Top Stories Past 30 Days

Most Popular

1. [Neural network AI is simple. So Stop pretending you are a genius](#)
2. [Top 10 Machine Learning Algorithms for Beginners](#)
3. [Comparing Machine Learning as a Service: Amazon, Microsoft Azure, Google Cloud AI](#)
4. [Want to Become a Data Scientist? Try Feynman Technique](#)
5. [Top 10 TED Talks for Data Scientists and Machine Learning Engineers](#)
6. [5 Fantastic Practical Machine Learning Resources](#)
7. [5 Machine Learning Projects You Should Not Overlook](#)

Most Shared

1. [Neural network AI is simple. So... Stop pretending you are a genius](#)
2. [5 Fantastic Practical Machine Learning Resources](#)
3. [A Simple Starter Guide to Build a Neural Network](#)
4. [Comparing Machine Learning as a Service: Amazon, Microsoft Azure, Google Cloud AI](#)
5. [Web Scraping Tutorial with Python: Tips and Tricks](#)
6. [Data Structures Related to Machine Learning Algorithms](#)
7. [A Beginners Guide to Data Engineering - Part I](#)

Latest News

- [A powerful new IDE to build, test, and run Apache Spark...](#)
- [Gartner 2018 Magic Quadrant for Data Science and Machin...](#)
- [The New Neural Internet is Coming](#)
- [Age of AI Conference 2018 - Day 2 Highlights](#)
- [Control Structures in R: Using If-Else Statements and L...](#)
- [Get the best out of your data - learn how, in Lon...](#)

More Recent Stories

- [Get the best out of your data - learn how, in London or ...](#)
- [University of Technology Sydney: PhD Research Studentships](#)
- [A Guide to Hiring Data Scientists](#)
- [5 Fantastic Practical Natural Language Processing Resources](#)
- [A Comparative Analysis of Top 6 BI and Data Visualization Tool...](#)
- [Top tweets, Feb 14-20: Neural Network AI is simple. Sg...](#)
- [Stanford online Data Science, Data Mining courses and certific...](#)
- [Age of AI Conference 2018 - Day 1 Highlights](#)
- [Recommender Engine- Under The Hood](#)
- [KDnuggets 18:n08, Feb 21: Neural network AI is simple-...](#)
- [Artificial Intelligence \(AI\) Conference, April 29 - May ...](#)
- [KDnuggets part-time, paid internship in Data Science/Machine L...](#)
- [PAW Vegas Early Bird Ends This Friday — Deep Learning and 4 ...](#)
- [Strata London, May 21-24 \(and Strata San Jose, Mar 5-8\) -...](#)
- [Top 20 Python AI and Machine Learning Open Source Projects](#)
- [Where AI is already rivaling humans](#)
- [Deep Learning Development with Google Colab, TensorFlow, Keras...](#)
- [Applied Data Science: Solving a Predictive Maintenance Busines...](#)
- [Deep Learning World Vegas - Talks from Cisco, Cap1, Lyft...](#)
- [5 Things You Need To Know About Data Science](#)

[KDnuggets Home](#) » [News](#) » [2017](#) » [Feb](#) » [Opinions, Interviews](#) » KDnuggets
Top Blogger: An Interview with Brandon Rohrer, Top Data Scientist ([17:n06](#))

© 2018 KDnuggets. [About KDnuggets](#)

[Subscribe](#): Email Name



X