Rounak Banik

Indian Institute of Technology Roorkee
rounakbanik@gmail.com
+91 84398 60325

Demystifying Data Science Conference at Metis

September 27, 2016

Overview

This report provides a short summary of my experience participating at the Demystifying Data Science Conference, an online meetup conducted by Metis on September 27, 2017 to get its participants acquainted and introduced to the world of Data Science.

Talks

The conference consisted of 30 minute talks by twenty four speakers over a span of 12 hours between 9AM and 9PM EST. I attended eight of the aforementioned talks and learnt a lot about the world of data science from renowned world experts on the subject. Topics covered included Statistics, Data Science Learning Resources, Networking, The Power of Social Media and Community, the Data Science Hiring Process, etc.

The first talk I attended was *5 Steps to Launch Your Data Science Career with Python* by Kevin Markham. The speaker explained the ever increasing influence of Python in the world of data science and why it is a very good investment to learn data science with Python. He then went on to give suggested learning path and major mistakes that beginners make (such as learning too much of Numpy but very little of Pandas).

This was followed by *Fixing the Hot Mess: What You're Doing Wrong to Hire and Get Hired in Data Science.* The speaker, Vin Vashista walked us through the typical tech stacks involved in data engineering and data science at various companies. He went on to explain the duality of strategy and productization and what companies and prospective employees must do to fix the complicated process of hiring in Data Science.

The next talk was a very insightful *Conversation with a Data Scientist* by Carla Gentry. She set to alleviate a lot of doubts and confusion regarding data science. She explained the difference between data science and big data, the importance of wrangling, the difference between analysts, statisticians and data scientists, etc.

Bob Hayes' talk on *The Practice of Data Science* applied data science to understand the world of data science. This meta talk presented the latest research on people, processes, technologies and organizations and what we can do better while looking out for jobs in the market.

Harnessing the Power of Community to excel in Data Science by Aylee Nielsen focused on the importance of having an active persona online. She explained the rising importance of platforms such as GitHub and Kaggle and how these platforms could be used to bring more attention to your work. The other parts of her talk focused on how to get people interested in your work internally and externally and also handed out very useful list of platforms to check ou to meet new Data Scientists.

The next talk on Statistics was by Deborah Berebichez. A Stanford Graduate, Deborah went on to explain how statistics is extremely misused in the world of data science and how we could take conscious steps towards exercising more caution in using statistical tools and gathering data in her talk entitled *Statistics and the Art of Deception.*

The final talk I attended was by Facebook Engineer Brandon Rohrer who spoke about *Open the Black Box: How Things Work Matters.* In his talk, Rohrer explained the need to deeply understand the machine learning algorithms you're applying to a particular problem. This helps in gaining a

better understanding of which algorithm would perform better in which cases. As an example, he demonstrated how he learnt about how SVMs work and how he would go about explaining the concept to a twelve year old.

Interaction

Since this was an online conference, the interaction with other participants was limited. However, I did manage to interact with a lot of budding data scientists in the Live Chat. I helped out a few people with their questions regarding learning Python, Statistics and Getting Started with Data Science. A few people on the forum also took a look at some of my work (Inferential Statistics Repository and the TED Talks Dataset) and I got a few follows, friend requests and connection requests on Twitter, Facebook and Linkedin respectively.