NYU Data Science Community features journalism, research papers, events, tools/software, and jobs for November 20, 2015.

Please let us (Laura Noren, Brad Stenger) know if you have something to add to the newsletter. We are grateful for the generous financial support from the Moore-Sloan Data Science Environment and to NYU's Center for Data Science.

NYU Data Science Community Newsletter Issue 028.

Data Science News

The Discovery of Statistical Regression

_Pricenomics, Dan Kopf_ from November 06, 2015

The legendary mathematician Carl Friedrich Gauss considered his alleged discovery of statistical regression “trivial.”

The method seemed so obvious to Gauss that he figured he must not have been the first to use it. He was sure enough it must have been discovered that he did not publicly state his finding until many years later, after his contemporary Adrien-Marie Legendre had published on the method. When Gauss suggested he had used it before Legendre it set off “one of the most famous priority disputes in the history of science...” Gauss would eventually be given most of the credit as the founder of regression, but not without a fight.

The Acute Pain and Chronic Reward of Public-Facing Work

_Trey Causey_ from November 17, 2015

Doing public-facing work, be it open source programming or writing, is an open invitation to allow internet commenters into your life. It can introduce anxiety, aggravate depression, induce anger, and more. There are certainly benefits and positives, but they are often more nebulous, revealing themselves over time in the form of "visibility" or "recognition". The criticisms manifest immediately and acutely.

A number of "I'm quitting my work on open source" posts have recently made the rounds on social media, with a common theme being that it's often quite literally thankless, that it consumes one's life without returning rewards, and demands that one work unpaid, on the weekends and at nights, to satisfy users... and for what?

Drug makers need to disclose more clinical trial data

_BMJ_ study by Jennifer Miller covered in STAT from November 12, 2015

35 percent of all trial results for 15 drugs that were approved by the Food and Drug Administration in 2012 were not disclosed. Additionally, nearly 30 percent of the trials conducted for those drugs failed to meet legal disclosure requirements.

“This confirms that pharmaceutical companies often fall below legal and ethical standards,” said Jennifer Miller, an assistant professor in the division of medical ethics at the New York University School of Medicine and a co-author of the study, which looked at 318 clinical trials involving almost 100,000 participants.

Machine Learning and Technical Debt with D. Sculley

_Software Engineering Daily_ from November 17, 2015
Technical debt, referring to the compounding cost of changes to software architecture, can be especially challenging in machine learning systems.

**D. Sculley** is a software engineer at Google, focusing on machine learning, data mining, and information retrieval. He recently co-authored the paper *Machine Learning: The High Interest Credit Card of Technical Debt.* [audio, 33:08]

**Computational Linguistics and Deep Learning**

*MIT Press Journals, Computational Linguistics, Christopher Manning* from September 25, 2015

Deep Learning waves have lapped at the shores of computational linguistics for several years now, but 2015 seems like the year when the full force of the tsunami hit the major Natural Language Processing (NLP) conferences. However, some pundits are predicting that the final damage will be even worse. Accompanying *ICML 2015* in Lille, France, there was another, almost as big, event: the 2015 Deep Learning Workshop. The workshop ended with a panel discussion, and at it, **Neil Lawrence** said, “NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.” Now that is a remark that the computational linguistics community has to take seriously!

**Toyota Research Institute Will Focus on Artificial Intelligence**

*IndustryWeek* from November 11, 2015

The **Toyota Research Institute Inc.** will straddle two coasts, with headquarters in Silicon Valley near Stanford University and a second facility in Cambridge, Mass., near the Massachusetts Institute of Technology. **Toyota** is also investing an additional $50 million each in **Stanford** and **MIT** to establish joint artificial intelligence research centers.

**Shrinking the accelerator**

*symmetry magazine* from November 19, 2015

The **Gordon and Betty Moore Foundation** has awarded $13.5 million to **Stanford University** for an international effort, including key contributions from the **Department of Energy’s SLAC National Accelerator Laboratory**, to build a particle accelerator the size of a shoebox. It’s based on an innovative technology known as “accelerator on a chip.”

This novel technique, which uses laser light to propel electrons through a series of artfully crafted glass chips, has the potential to revolutionize science, medicine and other fields by dramatically shrinking the size and cost of particle accelerators.

**“Shrinking bull’s-eye” algorithm speeds up complex modeling from days to hours**

*MIT News* from November 16, 2015

**MIT** researchers (**Youssef Marzouk, Patrick Conrad, Natesh Pillai**) have developed a new algorithm that vastly reduces the computation of virtually any computational model.

**NYC's Booming University Incubator Scene is Driving Innovation**

Observer; Sage Lazzaro, John Bonazzo and Brady Dale from November 15, 2015

Whether seeking prestige or profit, New York's top universities are schooling new entrepreneurs in the art of growing a company.

**rOpenSci Announces $2.9M Award from the Helmsley Charitable Trust**

Software Carpentry: Karthik Ram from November 13, 2015
rOpenSci, whose mission is to develop and maintain sustainable software tools that allow researchers to access, visualize, document, and publish open data on the Web, is pleased to announce that it has been awarded a grant of nearly $2.9 million over three years from The Leona M. and Harry B. Helmsley Charitable Trust. The grant, which was awarded through the Trust’s Biomedical Research Infrastructure Program, will be used to expand rOpenSci’s mission of developing tools and community around open data and reproducible research practices.

Why we need to create careers for research software engineers
Scientific Computing World, Simon Hettrick from November 11, 2015
Perversely, academic research penalises expert software developers for providing the tools on which the research relies. Time for a change, says Simon Hettrick of the Software Sustainability Institute.

Events
Data, Intuition, Models and Theories - Economics and Big Data Meetup
We are very pleased to host Professor Emanuel Derman, Columbia University. He directs their program in financial engineering. He is the author of My Life As A Quant. His latest book is Models.Behaving.Badly: Why Confusing Illusion with Reality Can Lead to Disasters, On Wall Street and in Life.

Friday, November 20, at 6 p.m., 251 Mercer St., Room 109

Is the Brain Bayesian? conference at NYU
Speakers and panelists will include: Jeffrey Bowers (Bristol), David Danks (Carnegie Mellon), Ernest Davis (NYU), Karl Friston (University College London), Weiji Ma (NYU), Larry Maloney (NYU), Eric Mandelbaum (CUNY), Gary Marcus (NYU), John Morrison (Barnard/Columbia), Nicoletta Orlandi (UC Santa Cruz), Michael Rescorla (UC Santa Barbara), Laura Schulz (MIT), Susanna Siegel (Harvard), Eero Simoncelli (NYU), Joshua Tenenbaum (MIT) and others

Friday-Saturday, December 4-5, at Kimmel Center and Jurow Hall, New York University

Sound Visualization & Data Sonification Hackathon
The Monthly Music Hackathon NYC for December.

Saturday, December 12, starting at 12 noon, Spotify, 45 W. 18th St.

Deadlines
KDD Cup 2016 Call for Proposals
We invite organizations' proposals for KDD Cup 2016. Starting in 1997, KDD Cup has been the most prestigious annual data mining competition held in conjunction with the ACM SIGKDD conference on Knowledge Discovery and Data Mining.

Deadline for Proposals is Monday, December 7.
Big Data special issue on Visualization in Data Science.
This Big Data journal special issue on Visualization in Data Science, expected to be published in March 2016, will showcase research in which data visualization is used to solve a data science problem.

Deadline for Manuscript Submissions: Friday, January 1, 2016

NYU/CNS-CDS: Faculty Position
The Center for Neural Science (CNS) and Center for Data Science (CDS) at New York University invite applications for an open rank faculty position from candidates with research programs aimed at quantitative analyses and modeling of neural data. We are particularly interested in scholars with cross-disciplinary research interests that complement those of our existing faculty.

The deadline for receipt of applications is Friday, January 15, 2016.

CDS News
Stern Professor Finds “Privacy-Friendly” Way to Identify Mobile Social Networks
NYU News from November 18, 2015
An analytical model that finds mobile device users based on their shared interests is a Holy Grail for e-marketers, because relevant, targeted campaigns have long been proven to be the most effective strategy with the highest return. Now NYU Stern Professor Foster Provost and co-authors have devised a way to connect the same and similar mobile users based on analyzing location visitation data – without compromising users’ privacy.

Click here to receive the NYU Data Science Community Newsletter OR to have us follow your twitter feed so that our data science twitter bot can easily grab links from your tweets.

To send us an announcement for the newsletter, please email laura.noren@nyu.edu and brad.stenger@nyu.edu by 9 pm Eastern Time on Thursday evenings for inclusion in Friday’s newsletter. We retain curatorial discretion.

NYU Data Science Community Newsletter Issue 028.