MICHELLE GILL, PH.D.

Data scientist, biophysicist

New York, NY

michelle@michellelynngill.com
203-901-4028

curriculum vitae and publications
mlgill
themodernscientist.com
michellelynngill
modernscientist

EDUCATION

Ph.D. Molecular Biophysics & Biochemistry
2006, Yale University, *New Haven, CT*

B.S. Biochemistry, *Summa Cum Laude*
2001, University of Kansas, *Lawrence, KS*

SKILLS

Programming: Python, Matlab, C, R, Shell

Big data: Spark, Dask

Machine learning: Supervised and unsupervised algorithms**,** Clustering, Natural language processing, Principal component analysis, Recommender systems, Computer vision, Neural networks, Deep learning, Compressed sensing

Statistics: Regression (linear, logistic, non-linear), Monte Carlo simulations

Databases: PostgreSQL, MongoDB

Other: Unix, Git, AWS, Docker, Flask, LaTeX

AWARDS

- Ruth L. Kirschstein National Research Service Postdoctoral Fellowship
- NSF Graduate Research Fellowship
- Barry M. Goldwater Scholar
- Kansas Board of Regents full-tuition merit scholarship

EXPERIENCE

Senior Data Scientist **2016 – Present**
Metis
- Designed and created Hadoop and Spark machine learning and NLP curriculum using self-made Docker containers
- Conducted corporate trainings focused on Python, Spark, Hadoop, and Hive
- Developed 12-week machine learning course for F100 company
- Co-instructed 12-week data science bootcamps
- Developed and conducted take home coding exercise to assist with interview preparation

Research Scientist **2014 – 2016**
National Cancer Institute
- Developed NESTA-NMR, which uses compressed sensing to enable up to 100X faster acquisition of large (~10 GB) experimental data sets
- Created website and documentation for NESTA-NMR
- Elucidated mechanisms of cancer pathway using high-resolution experimental techniques

Postdoctoral Research Fellow **2008 – 2014**
Columbia University Medical Center
- Used Monte Carlo simulations to model effect of physical changes on enzyme activity
- Developed MFOutParser, a Python library that parses a challenging text format, enabling 10X faster analysis times
- Member of team that studied anti-oncogenic associated mechanisms of substrate binding to AlkB, a DNA repair enzyme

Postdoctoral Research Fellow **2007 – 2008**
University of Kansas
- Developed principal component analysis-based method to visualize changes in vaccine structure that are critical for efficacy

Consultant **2006 – 2007**
The Boston Consulting Group
- Developed Excel-based statistical tools and Access database for organizational streamlining of pharmaceutical client
- Part of team that developed municipal bond investment strategy for financial services client
- Member of team awarded 2007 Global Strategy Olympics Prize for pharmaceutical client work

PRESENTATIONS & PROJECTS

- "Learning from Text: Natural Language Processing with Python", Tutorial, *ODSC East*, Boston
- Created wine label recognition application computer vision
- pdLSR is a library for performing linear and non-linear least squares regression in a dataframe-aware fashion