Steve Herrin

CONTACT INFORMATION

jobs@steveherrin.com
650-814-8865
San Jose, CA

www.github.com/steveherrin
www.linkedin.com/in/herrinsteve
medium.com/@steveherrin

EXPERIENCE

23andMe, Sunnyvale, CA — January 2014 – present

Software Engineering, Machine Learning, Data (titles: sr. software engineer, tech lead) (4.5 yrs)

- Developed and automated a maximum likelihood analysis that flagged $\sim 1\%$ bad genotyping probes for replacement in future platforms
- Automated genotype calling using a combination of unsupervised and supervised ML techniques
- Architected and led building of systems to ensure quality and reproducibility of machine learning models, used for over 90% of production models
- Designed and implemented a Python library providing a unified API for accessing customer data across MySQL, HBase, and other data stores, simplifying building and deploying machine learning models
- Built 3 generations of data pipelines with Celery, Luigi, and AWS Simple Workflow to run Python, C++, and R algorithms that impute, transform, and analyze petabytes of genetic data
- Led team building a web-based portal that allows internal and external researchers to dynamically query anonymized data stored in HBase for >5 million customers
- Created Genotyping Services, a Django webapp on AWS allowing external researchers to easily run genomic studies, increasing sales by over 2% and leading to strategic data-sharing agreements

People Management (title: engineering manager) (1.5 yrs)

- Built 3 data-focused engineering teams totaling 16 engineers, including a mix of leads, senior, and junior level individual contributors
- Led team of engineering leads to standardized interviewing guidelines and open source release processes

SLAC National Accelerator Lab, Menlo Park, CA — May 2008 – August 2013

Research Associate

- Applied machine learning and computer vision algorithms to improve detector energy resolution by 25%
- Repurposed the detector for 3D cosmic ray muon reconstruction, yielding a 10x reduction in cosmogenic background uncertainty
- Built and maintained slow control systems collecting heterogeneous sensor data at 1 Hz from over 600 channels at a site with unreliable internet connectivity to protect $10M of liquid xenon
- Developed batch data pipelines using Python, C++, and shell scripts to process TB of calibration data

Rice University, Houston, TX — May 2005 – May 2007
and **University of Washington**, Seattle, WA — June 2006 – August 2006

Undergraduate Research Assistant

- Implemented (in C++) and evaluated random forest and boosted decision tree algorithms that contributed to the discovery of single top quark production by Fermilab's D0 experiment

SKILLS

Languages: Python, Rust, Elm, C++, SQL, Shell Scripting, JavaScript/TypeScript, MATLAB/Octave, R

Tools: AWS, NumPy, SciPy, Scikit-Learn, Django, MySQL, PostgreSQL, Git, LaTeX, HBase, Spark

Other: Machine Learning, Data Analysis, Statistics, Simulation, Sensors, Neutrino & Particle Physics

OPEN SOURCE

SpookyOTP: A lightweight Python implementation of TOTP/HOTP authentication

EDUCATION

Insight Data Science, Mountain View, CA — December 2013
- Postdoctoral Fellowship

Stanford University, Stanford, CA — June 2013
- Ph.D. (Physics)

Rice University, Houston, TX — May 2007
- B.S. (Physics)