# 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**, Mountain View, CA

*Engineering Manager: Machine Learning & Data Services Team*      March 2018 – present

- Proposed and built a team to develop common machine learning infrastructure, freeing up 5 other science and engineering teams from having to maintain their own
- Authored engineering's open source policy, streamlining the release of 2 (to date) open source packages
- Worked with engineer leads to develop standard interviewing guidelines across the organization

**23andMe**, Mountain View, CA

*Software Engineering Individual Contributor (final title: Tech Lead)*      January 2014 – March 2018

- Architected systems to ensure quality and reproducibility of machine learning models, now used for over 90% of production models
- Created Genotyping Services, a Django webapp on AWS allowing external researchers to easily run genetic studies, increasing sales by over 2% and leading to strategic data-sharing agreements
- Led the building of a web-based portal that allows internal and external researchers to dynamically query anonymized data stored in HBase for >5 million customers
- Designed and implemented a Python library that provides a unified API for accessing customer data across MySQL, HBase, and other data stores
- 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

**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%
- Created a PHP logbook webapp with a MySQL backend for tracking work on the EXO-200 experiment
- Collected heterogeneous sensor data at 1 Hz from over 600 channels at a site with unreliable internet connectivity to protect $10M of liquid xenon and for use in later analysis
- Developed batch data pipelines using Python, C++, and shell scripts to process TB of calibration data
- Mentored 1–2 junior graduate students (at any given time) on lab, coding, and statistical technique

**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, Celery, Luigi, 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

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

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