SEAN HARNETT

srharnett@gmail.com 646-942-1043 www.seanharnett.me github.com/sharnett

Experience

Google
New York, New York

Software Engineer, Research Feb 2016 - present

- Work on Coauthor, a service which dynamically generates cross-lingual multi-modal content using many of Google's sources of richly annotated data, as seen in Google Docs Explore (millions of monthly active users)
- Create and run large distributed computing jobs (C++ in Flume, MapReduce) using thousands of machines on datasets up to one petabyte
- Designed and implemented a fast automated quality evaluation framework which enabled rapid improvements in quality, ultimately allowing for an initial launch in twelve languages. Created (mostly Python) dashboards, monitoring, and alerts to track changes in performance and prevent regressions

Facebook
Menlo Park, California

Data Scientist, Analytics Intern Jun 2015 - Aug 2015

- Leveraged big data tools (Hadoop, Hive, Presto) to drive product decisions for Facebook's lightweight advertising products, impacting millions of advertisers
- Designed and analyzed large-scale experiments on budget defaults, discovering actionable insights leading to significant improvements in revenue and number of active advertisers

Yodle
New York, New York

Data Scientist Jun 2013 - Apr 2015

- Developed statistical models and algorithms for paid advertising bidding strategies for tens of thousands of small business clients and millions of dollars of annual ad spend
- Scraped logs with grep/sed/awk/etc., wrote complex SQL queries, explored and prototyped models in R and Python, implemented production systems in Java, ran and evaluated dozens of live experiments, visualized and communicated results to non-technical stakeholders
- Designed experiments and analyzed results for A/B testing of thousands of clients' websites
- Architected and implemented fully-automated customer phone call transcription and classification pipeline, including pulling and processing data, submitting to third-party transcription API, training and evaluating classification models, assembling predictions for business applications (identifying booked appointments, spam and wrong number detection)

Los Alamos National Laboratory, Center for Nonlinear Studies
Los Alamos, New Mexico

Graduate Student Researcher Jun-Jul 2011, May-Jul 2012

Various English Schools
Daegu, South Korea; Brighton, UK; Barcelona, Spain

English Teacher Apr 2008 - Jun 2009

Entertainment Brokers International
Los Angeles, California

Systems Designer Sep 2006 - Feb 2008

Education

Columbia University
PhD in Applied Mathematics, February 2016
Dissertation title: *Optimization methods for power grid reliability*
Developed computational tools for power grid reliability using Python, MATLAB, Gurobi, CPLEX.

University of California, Los Angeles
MA in Mathematics, 2006
BS in Applied Mathematics with Specialization in Computing, 2006

Skills

Languages: Python, C/C++, SQL, R, MATLAB, Java, Julia

Tools: IPython notebook, pandas, scikit-learn, RStudio, ggplot2, dplyr, vim, git, vowpal wabbit, Stan

Areas of interest: numerical algorithms, machine learning, optimization, data visualization