

Scott Clark

Data Mining Engineer and Team Lead at Yelp
<http://github.com/sc932>

April 14, 2016
scott@scottclark.io
<http://scottclark.io>

Education

- **Cornell University** Ithaca, NY
Ph.D. Applied Math, M.S. Computer Science 2008 - 2012
 - Department of Energy Computational Science Graduate Fellow (Full Scholarship, 4 years)
 - Emphasis on machine learning/data mining and algorithm design/software development related to bioinformatics and optimization
 - Committee: Prof. Peter Frazier (advisor), Prof. Steve Strogatz, Prof. Bart Selman, Dr. Zhong Wang
- **Oregon State University** Corvallis, OR
B.Sc. Mathematics, B.Sc. Computational Physics, B.Sc. Physics 2004 - 2008
 - Graduated Magna Cum Laude with minors in Actuarial Sciences and Mathematical Sciences
 - Strong emphasis on scientific computing, numerical analysis and software development
 - Advisors: Prof. Rubin Landau, Prof. Malgorzata Peszynska

Industry Experience

- **SigOpt Inc - <https://sigopt.com>** San Francisco, CA
Co-founder and CEO November 2014 - current
 - Helping lead a world class team to eliminate expensive trial and error from every experts workflow. Using cutting edge optimization behind a simple API to help tune machine learning models and build better products in a variety of fields.
- **Yelp Inc** San Francisco, CA
Data Mining Engineer and Lead on Ad Targeting Team July 2012 - current
 - **Optimization:** Co-developed and led team for MOE: the Metric Optimization Engine (github.com/Yelp/MOE, an **open source** optimization framework), found significant gains in different metrics across the organization using Bayesian Global Optimization algorithms.
 - **Targeting:** Implemented multi-armed bandit strategies for ad selection, sole targeting engineer on mobile app ads rollout, developed new location-based targeting algorithms, advised and helped develop other machine learning and math based targeting projects.
 - **Leadership:** Led cross-organization teams as director of the Yelp Dataset Challenge, MOE team lead, intern and new-hire mentor, founder and organizer of Applied Learning Group (bi-weekly, all-engineering technical speaker series) and more. Successfully managed projects from inception to on-time completion; balancing tradeoffs and coordinating teamwork.
 - **Recruiting:** Created, implemented, and directed yelp.com/dataset_challenge, gave many dozens of tech talks across the country, led events, gave hundreds of technical interviews, and closed candidates.
- **Bloomberg LP** New York, NY
Financial Software Development Intern Summer 2011
 - Developed end-to-end reporting software in C++ and javascript

- Implemented statistical models to perform forward and backward portfolio analysis

Research Experience

- **DOE Joint Genome Institute (Lawrence Berkeley National Lab)** Walnut Creek, CA
Researcher in Analysis Group under Dr. Zhong Wang Summer 2010
 - Created **open source** genome validation software tool in **python** and **C**
 - Used machine learning to mine TBs of genome data efficiently using novel likelihood function
- **Los Alamos National Laboratory** Los Alamos, NM
Researcher in Metagenomics Group under Dr. Nick Hengartner Summer 2009
 - Wrote **open source** alignment algorithm software tool in **python**, **C** and **CUDA**
 - Used statistical models to discover sequence alignments using parallel algorithms on GPUs
- **Oregon State University** Corvallis, OR
Research Assistant under Prof. Malgorzata Peszynska and Prof. Rubin Landau 2005-2008
 - Finite element analysis with uncertainty and web-based teaching in **Java**
- **Max Plank Institute for the Physics of Complex Systems** Dresden, Germany
NSF REU Research Assistant under Prof. Steven Tomsovic Summer 2007
 - Research on extreme value statistics in **MATLAB** and **FORTRAN**
- **University of California: Davis** Davis, CA
NSF REU Research Assistant under Prof. Daniel Cox Summer 2006
 - Computational biophysics research as applied to protein folding in **Java**

Publications

1. SC Clark, *PARALLEL MACHINE LEARNING ALGORITHMS IN BIOINFORMATICS AND GLOBAL OPTIMIZATION*. PhD Thesis in Applied Mathematics, Cornell University, May 2012.
2. SC Clark, R Egan, P Frazier, and Z Wang, *ALE: an Assembly Likelihood Evaluation Framework to Assess the Accuracy of Metagenome Assemblies* *Bioinformatics* **29**, 4, 435-443 (2013).
3. SC Clark, N Hengartner, and J Berendzen, *Velvetropo: a parallel, bitwise algorithm for nding homologous regions within multiple sequences* submitted to *BMC Bioinformatics* (2010).
4. SC Clark, *Solving Genomic Jigsaws*. *DEIXIS Magazine* **8** 30–32 (2010).
5. KC Kunes, SC Clark, DL Cox, and RR Singh, *Left handed helix models for mammalian prion fibrils*. *Prion* **2**, 2, 81–90 (2008).

Presentations

- *Introducing the Metric Optimization Engine (MOE); an open source, black box, Bayesian Global Optimization engine for optimal experimental design*
 - MLconf 2014 (Talk) – San Francisco, CA (October 2014)
 - Stanford University AI Lab (Talk) – Palo Alto, CA (October 2014)

- San Francisco Machine Learning Meetup (Talk) – San Francisco, CA (October 2014)
- ACM RecSys 2014 (Talk) – Palo Alto, CA (October 2014)
- Stanford University (Talk) – Palo Alto, CA (October 2014)
- Microsoft Research (Talk) – Mountain View, CA (September 2014)
- University of California, Berkeley (Talk) – Berkeley, CA (September 2014)
- Cornell University (Talk) – Ithaca, NY (September 2014)
- Hack Lab (Talk) – Zagreb, Croatia (September 2014)
- BounceExchange (Talk) – New York, NY (August 2014)
- BrightRoll (Talk) – San Francisco, CA (August 2014)
- LinkedIn (Talk) – Mountain View, CA (August 2014)
- Netflix (Talk) – Los Gatos, CA (August 2014)
- Optimizely (Talk) – San Francisco, CA (August 2014)
- *Optimal Learning for Fun and Profit, an introduction to multi-armed bandits*
 - Yelp Headquarters (Several Talks) – San Francisco, CA (2013-2014)
 - Cornell University (Talk) – Ithaca, NY (February 2014)
 - Stanford University (Talk) – Palo Alto, CA (January 2014)
 - University of California, Davis (Talk) – Davis, CA (May 2014)
 - Cornell University (Talk) – Ithaca, NY (September 2013)
 - Carnegie Mellon University (Talk) – Mountain View, CA (April 2013)
 - Harvey Mudd College (Talk) – Claremont, CA (April 2013)
 - Cornell University (Talk) – Ithaca, NY (February 2013)
 - Stanford University (Talk) – Palo Alto, CA (January 2013)
 - Cornell University (Talk) – Ithaca, NY (September 2012)
- *Parallel Machine Learning Algorithms in Bioinformatics and Global Optimization, Thesis Defense*
 - DOE CSGF Annual Conference (Talk) – Washington D.C. (June 2012)
 - Cornell University (Talk) – Ithaca, NY (May 2012)
- *Learning for Metagenomic Assembly Validation and Optimization*
 - Supercomputing 2011 (SC11) (Poster) – Seattle, WA (November 2011)
 - DOE CSGF Annual Conference (Poster, award finalist) – Washington D.C. (June 2011)
 - SIAM CSE conference (Talk) – Reno, NV (February 2011)
 - INFORMS annual meeting (Talk) – Austin, TX (November 2010)
 - Cornell Math Sciences Seminar (Talk) – Ithaca, NY (November 2010)
 - DOE Joint Genome Institute Visiting Speaker (Talk) – Walnut Creek, CA (August 2010)
- *Development and Exploration of Velvetrope: a bitwise, parallel alignment algorithm on biological sequences*
 - Supercomputing 2010 (SC10) (Poster, ACM student competition) – New Orleans, LA (November 2010)
 - Cornell Math Sciences Seminar (Talk) – Ithaca, NY (November 2010)

- DOE CSGF Annual Conference (Poster, award finalist) – Washington D.C. (June 2010)
- *Redundancy in random k -SAT*
 - Cornell Engineering Research Conference (Poster) – Ithaca, NY (May 2010)
 - DOE CSGF Annual Conference (Poster, award finalist) – Washington D.C. (June 2009)
- *Finite Element Analysis of Uncertain Interfaces*
 - Oregon State University Computational Mathematics Seminar (Talk) – Corvallis, OR (June 2008)

Recruiting Events

- Carnegie Mellon University
 - Seminar Talk (speaker) – April 2013
- Cornell University
 - Career Fair – September 2014
 - Tech Networking Reception – September 2014
 - Resume Review and Office Hours – September 2014
 - Tech Talk (speaker) – September 2014
 - On campus interviews – September 2014
 - Career Fair – February 2014
 - Tech Networking Reception – February 2014
 - Resume Review and Office Hours – February 2014
 - Tech Talk (speaker) – February 2014
 - On campus interviews – February 2014
 - Career Fair – September 2013
 - Resume Review and Office Hours – September 2013
 - Tech Talk (speaker) – September 2013
 - On campus interviews – September 2013
 - Career Fair – February 2013
 - Tech Talk (speaker) – February 2013
 - Resume Review and Office Hours – February 2013
 - On campus interviews – February 2013
 - Career Fair – September 2012
 - Tech Talk (speaker) – September 2012
 - On campus interviews – September 2012
- Harvey Mudd College
 - Tech Talk (speaker) – April 2013
 - 5C Hackathon (speaker, judge) – April 2013

- Insight Data Science
 - Data Science Fellows Presentations – October 2014
 - Data Engineering Fellows Presentations – October 2014
 - Data Engineering Fellows Pitch (speaker) – September 2014
 - Data Engineering Fellows Presentations – July 2014
 - Data Science Fellows Presentations – July 2014
 - Data Engineering Fellows Pitch (speaker) – July 2014
 - Data Science Fellows Presentations – February 2014
 - Data Science Fellows Pitch (speaker) – January 2014
 - Data Science Fellows Presentations – October 2013
 - Data Science Fellows Presentations – July 2013
 - Data Science Fellows Pitch (speaker) – June 2013
 - Data Science Fellows Pitch (speaker) – February 2013
- Stanford University
 - Tech Talk (speaker) – October 2014
 - Big Hack (speaker, judge) – April 2014
 - Stanford AI Lab Masters Presentations (speaker) – March 2014
 - Tech Talk (speaker) – January 2014
 - Tech Talk (speaker) – January 2013
- University of California, San Diego
 - Career Fair – October 2013
 - Tech Talk (speaker) – October 2013
 - On Campus Interviews – October 2013
- University of California, Davis
 - Mathematics Department Seminar (speaker) – May 2014
 - Mathematics Seminar Reception and Dinner – May 2014
- University of California, Berkeley
 - Tech Talk (speaker) – September 2014
 - Information School Final Project Presentations (judge) – May 2014
- Yelp Headquarters
 - Dozens of closing calls and dinners – 2012-2014
 - 200+ Software Engineering and Data Science interviews – 2012-2014
 - San Francisco Data Science Meetup (introductory speaker) – August 2014
 - Intern NITE (speaker) – July 2014
 - Berkeley @ Yelp Tech Talk (speaker) – April 2014
 - Yelp Open House Event (speaker) – November 2013
 - Intern NITE (speaker) – July 2013

- Stanford LifeSwap (speaker) – January 2013
- Intern NITE (speaker) – October 2012

Participation in Workshops and Conferences

- **KDD 2014** New York, NY
Association for Computing Machinery/SIGKDD August 24-27, 2014
- **WWW 2013** Rio de Janero, Brazil
World Wide Web Conference May 13-17, 2013
- **Supercomuputing 2011** Seattle, WA
Association for Computing Machinery/IEEE November 12-18, 2011
- **SIAM Conference on Computational Science and Engineering** Reno, NV
Society for Industrial and Applied Mathematics Feb 28-March 4, 2011
- **Supercomuputing 2010** New Orleans, LA
Association for Computing Machinery/IEEE November 13-19, 2010
- **INFORMS Annual Meeting 2010** Austin, TX
Institute for Operations Research and the Management Sciences November 7-10, 2010
- **ICCS Computational Science Summer School (Many-core)** Oakland, CA
International Center for Computational Science August 2-6, 2010
- **Computational Science Graduate Fellowship Annual Conference** Washington D.C.
Krell Institute June 22-24, 2010
- **High Performance Computing Workshop** Washington D.C.
Krell Institute June 21, 2010
- **Cornell Engineering Research Conference** Ithaca, NY
Cornell College of Engineering March 16, 2010
- **Supercomputing 2009** Portland, OR
Association for Computing Machinery/IEEE November 14-20, 2009
- **q-bio Conference on Cellular Information Processing,** Los Alamos/Sante Fe, NM
Los Alamos National Laboratory Center for Non-Linear Studies August 5-9, 2009
- **q-bio Summer School on Cellular Information Processing** Los Alamos/Sante Fe, NM
Los Alamos National Laboratory Center for Non-Linear Studies July 20 - August 4, 2009
- **Computational Science Graduate Fellowship Annual Conference** Washington D.C.
Krell Institute July 14-16, 2009
- **High Performance Computing Workshop** Washington D.C.
Krell Institute July 13, 2009
- **DOE ACTS Workshop (Computational Science Tools)** Berkeley, CA
NERSC/LBNL/ASCR/DOE Office of Science August 19-22, 2008
- **Computational Science Graduate Fellowship Annual Conference** Washington D.C.
Krell Institute June 16-19, 2008
- **Risk Analysis: Perception, Policy and Practice Workshop** Research Triangle Park, NC
Statistical and Applied Mathematical Sciences Institute October 3-4, 2008
- **Kickoff Workshop of the SAMSI program on Risk Analysis** Research Triangle Park, NC
Statistical and Applied Mathematical Sciences Institute September 16-19, 2007
- **SAMSI/CRSC Undergraduate Modeling Workshop** Raleigh, NC
SAMSI/NCSSU/Duke May 21-25, 2007

Awards, Grants & Honors

Forbes 30 Under 30, Enterprise Tech	2016
Department of Energy Computational Science Graduate Fellowship (CSGF) (\$300 000)	2008-2012
NERSC Production Allocation (PI) (100 000 Cray XT4 hours)	2012
NERSC Startup Allocation Renewal (PI) (15 000 Cray XT4 hours)	2012
Cornell University Conference Travel Grant (\$390)	2011
NERSC Startup Allocation Renewal (PI) (15 000 Cray XT4 hours)	2011
Cornell University Conference Travel Grant (\$390)	2010
NERSC Startup Allocation (PI) (15 000 Cray XT4 hours)	2010
DOE CSGF Essay Contest Honorable Mention (\$500)	2010
Cornell University Sage Fellowship (\$55 000, declined)	2008-2009
Joel Davis Award in Mathematics (\$1 000)	2007-2008
URISC Undergraduate Research Fellowship (\$1 500)	2007-2008
NSF Research Experience for Undergraduates Program (MPI PKS) (\$6 000)	2007
Paul Copson Memorial Scholarship in Physics (\$1 000)	2006-2007
NSF Research Experience for Undergraduates Program (UC Davis) (\$6 000)	2006
Nicodemus Scholarship in Physics (\$1 000)	2005-2006
Diversity Achievement Scholarship (\$4 000)	2004-2008

Selected Open Source Projects (github.com/sc932)

- **MOE: Metric Optimization Engine** Python, C++, CUDA
Global, black box optimization of expensive functions 2012 - Current
- **ALE: Assembly Likelihood Estimator** C, Python
Probabalistic evaluation of genome assemblies 2010 - 2012
- **Velvetrope** Python, C, CUDA
A parallel statistical algorithm for finding homologous regions within sequences 2009 - 2010
- **BetaHelix** Java
Computes various statistics about a left or right handed beta helix 2006 - 2007

Skills

- **Numerical Analysis and Computer Science:** Machine Learning, Data Mining, Optimization, Computational Science, Artificial Intelligence, Linear Algebra, Monte Carlo Methods, ODEs, PDEs, Iterative Methods, Parallel Programming, Distributed Systems, Data Structures
- **Public Speaking:** I've given several hundred technical talks to audiences at machine learning conferences, Fortune 500 boards, and beyond.
- **Development:** Python (preferred), C/C++, CUDA, JavaScript, L^AT_EX
- **Technology:** MapReduce, numerical libraries, web frameworks, databases, linux, git, vim, tmux
- Discovering and implementing new ideas. Give me an API and a problem and I will figure it out.

- Diverse background in Math, Computer Science, Physics and Biology allows me to communicate to a wide scientific and general audience and begin contributing to any group immediately.
- I have worked in many places in a myriad of fields. I can readily learn and adapt to a new discipline, area or environment and start pushing real results quickly.