

Sam Steingold

sds@podval.org

<http://sds.podval.org/resume.html>

Summary: PhD Mathematician/Data Scientist with extensive expertise in machine learning; software and algorithm design and development; team building and management experience, and more.

Skills:

Applied Mathematics — mathematical modeling using information theory, graph/social network theory, machine learning, free text analysis, Bayesian modeling, stochastic processes, regression analysis, neural networks, support vector machines.

Software Development — designed and developed numerous software packages.

Operating Systems: UNIX (Linux, Mac OS X, Solaris), Windows.

Programming Languages: Python; R; C/C++/Java; Scala, OCaml; Lisp/Scheme; Perl; Pig, Hive, SQL, SAS; Matlab, Maple, Mathematica, Maxima etc.

Distributed Data Processing: Hadoop, Spark.

Research and Presentation — conducted independent and joint research, wrote papers and gave talks on the results.

Financial Markets — experience with trading modeling and systems design and development; knowledge of several derivatives pricing methods, such as stochastic differential equations (Black-Scholes analysis), risk-neutral valuation, binomial trees; hedging strategies.

Management — built and managed a Data Science team of local and remote permanent members and contractors; mentored junior staff (machine learning and software development); managed half a dozen contributors to a large Open Source project.

Employment history:

- **CLEAR** (New York, NY), Chief Data Scientist August 2016 – present
CLEAR does instant biometric identification
- **Magnetic Media Online** (New York, NY), Lead Data Scientist October 2013 – July 2016
Magnetic is an Internet marketing company specializing in search retargeting
Grew and managed the Data Science team.
Mentored user response modeling and fraud detection efforts.
Designed a search-history-based marketing campaign optimization machine learning system.
Developed a look-alike audience modeling methodology.
Spearheaded the data provider quality monitoring effort.
Contributed to the query categorization research.
Investigated the impact of observed user behavior on campaign response.
Algorithms implemented in **R, Python, Scala, Pig, HiveSQL**.
- **AddThis** (New York, NY), Data Scientist November 2011 – October 2013
AddThis is an Internet marketing company
Designed and implemented a text sentiment analysis engine.
Built look-alike models for audience optimization.
Analyzed a social network arising from URL sharing.
Analyzed user browsing patterns (clustering, correlations).
Algorithms implemented in **R, Perl, Java**.
- **ALGO Engineering** (New York, NY), Quantitative Developer January 2011 – October 2011
ALGO Engineering is a trading technology company
Designed and implemented new trading strategies; optimized performance of existing strategies.
Algorithms implemented in **C++**.
- **Jane Street Capital** (New York, NY), Quantitative Researcher February 2006 – November 2010
Jane Street is a hedge fund

Developed and analysed various statistical arbitrage, trading and hedging strategies using regression analysis, Bayesian algorithm, and text analytics.

Developed various tools to support research, development, and trading.

Mentored junior developers and interns.

Algorithms implemented in **OCaml**.

- **BAE Systems AIT** (Burlington, MA), Lead Research Engineer June 2003 – January 2006

BAE Systems AIT – formerly ALPHATECH – is a leader in developing sophisticated mathematical and knowledge-based algorithms

Conducted research on group discovery and vertex classification in large simulated social networks.

Developed adaptive computer network protocol optimization.

Algorithms implemented on **Common Lisp** and **C++**.

- **Premonitia, Inc** (Acton, MA), Principal Mathematical Engineer April 2002 – October 2002

Premonitia tried to predict network faults before they actually happen, until it shut down

Developed the methodology to identify the kind of fault being predicted (system crash versus link down etc) using Bayesian learning and support vector machines.

Algorithms implemented in **Matlab**, **C/C++**.

- **Xchange, Inc** (Boston, MA), Senior Analyst July 1999 – February 2002

Xchange was a leading CVM (Customer Value Management) software provider, until it shut down

Took an active part in consulting and pre-sale engagements; built predictive models for cross-sell and attrition using neural networks and Bayesian learning; worked closely with the clients on problem specification.

Mentored junior staff on modeling and software development.

Implemented a Bayesian learning algorithm with boosting and many other data analysis, cleaning, processing and transformation tools in **Java**, **Perl**, **Common Lisp**, **SAS** and **C**.

Participated in the **PMML** standardization process with the **DMG**.

- **Eagle Trading Systems** (Princeton, NJ), Quantitative Analyst July 1997 – June 1999

Eagle is a small futures trading company, a CTA (Commodities Trading Advisor)

Designed and developed technical futures trading systems and hedging strategies based on pattern recognition (channel identification). Conducted performance analysis for a market-neutralized futures trading system.

Algorithms implemented in **Common Lisp**.

- **Parametric Technology Corp** (Waltham, MA), Software Engineer July 1996 – June 1997

PTC is a CAD/CAM software industry leader

Developed various geometric creation and analysis algorithms and implemented them in **C** in a cross-platform environment, which included all major UNIX variants as well as Windows NT.

- **UCLA Mathematics Department** (Los Angeles, CA), Sloan Research Fellow 1995 – 1996

Conducted research in the area of differential geometry, metric invariants of Alexandrov spaces, and comparison geometry with various curvature bounds.

Algorithms implemented in **Mathematica**, **Matlab** and **Maple**.

- **UCLA Mathematics Department** (Los Angeles, CA), Research and Teaching Assistant 1992 – 1995

Taught lower and upper division classes (Differential Equations, Differential Geometry, Analysis)

- **Manufacturing Research Corporation** (Moscow, Russia), Research Analyst 1991 – 1992

MRC was a leading Russian car industry research institution

Developed a statistical method for analysis of the dependence of the car parts' precision quality on the manufacturing process, material and other 14 factors. Developed a software system implementing the statistical analysis method and evaluated the technical standards used in car manufacturing.

Education:

- **University of California** (Los Angeles, CA), Department of Mathematics Sept 1992 – June 1996

GPA: 4.0. **MA** in Mathematics – June 1993, **PhD** in Mathematics – June 1996.

- **Moscow Aviation Institute** (Russia), Department of Applied Math and CS Sept 1986 – Feb 1992

GPA: 3.9. **BS/MS** in Applied Mathematics and Computer Science – Feb 1992.

Achievements: The list of publications, conferences, languages, and software attached.

Sam Steingold - attachment

Honors:

- Five time winner of the Moscow Aviation Institute Student Mathematical Competition (1987-91).
- First Prize, Advanced Probability/Statistics Competition, Moscow Aviation Institute, 1990.
- Second Prize, Moscow City Student Mathematical Competition, 1989.
- Member, MAA (Mathematical Association of America).
- Member, ACM SIGKDD.
- Alfred P. Sloan Dissertation Year Fellowship 1995/96 (one of the most prestigious national fellowships in mathematics).

Publications:

1. S. Steingold, D. Crosta, B. Leung, J. Arfa "Monotonically Increasing Multi-Interval Continuous Calibration", *submitted to CIKM 2016*, <https://github.com/Magnetic/MIMIC>.
2. S. Steingold, M. Laclavík "An Information Theoretic Metric for Multi-Class Categorization", *MLConf 2016*, <https://github.com/Magnetic/proficiency-metric>.
3. M. Laclavík, S. Steingold and M. Ciglan "Detecting Brands in User Search Queries", *WSDM QRUMS 2016 Workshop*, 2016.
4. M. Laclavík, M. Ciglan, A. Dorman, S. Dlugolinsky, S. Steingold, M. Šeleng. "A Search Based Approach to Entity Recognition: Magnetic and IISAS team at ERD Challenge", *ERD2014*, Gold Coast, Queensland, Australia, July 2014.
5. S. Steingold, C. Fournelle, J.V. White. "Clustering and Threat Detection", *2005 AAAI Spring Symposium on AI Technologies for Homeland Security*, Stanford University, March, 2005.
6. J.V. White, S. Steingold, C. Fournelle. "Performance Metrics for Group-Detection Algorithms", *Interface 2004*, Baltimore, MD, May 29, 2004.
7. J.V. White, G. Colby, S. Steingold. "Threat-Detection Performance On Wet-Run Data", *EAGLE miniTIE workshop*, August 2, 2004.
8. Sam Steingold, Richard Wherry, Gregory Piatetsky-Shapiro, "Measuring Real-Time Predictive Models", *ICDM2001*.
9. Gregory Piatetsky-Shapiro, Sam Steingold, "Measuring Lift Quality in Database Marketing", *SIGKDD Explorations*, Vol. 2:2, (2000), 81-86.
10. Peter Petersen, Semion Shteingold, Guofang Wei, "Comparison Geometry with Integral Curvature Bounds", *GAGA* 7 (1997) 1011-1030.
11. Semion Shteingold, "Covering Radii and Paving Diameters of Alexandrov Spaces", *J. of Geometric Analysis*, 8 (1999), no. 4, 613-627.
12. Semion Shteingold, "Volume comparison for Alexandrov spaces", *Indiana U. Math. J.*, 43 (1994), no. 4, 1349-1357.
13. Semion Shteingold, "One Solution of the Pentagon Problem", *J. of Math. Behavior*, 13 (1994), 299-302.
14. Semion Shteingold, "Smoothness of Generalized Solutions of Boundary Value Problem for Functional-Differential Equations", *Uspehi Mat. Nauk (in Russian)*, 46 (1991), no. 3(279), 203-204.

Conferences Attended:

- Micro Program in Riemannian Geometry; Fields Institute for Research in Mathematical Sciences, Ontario, Canada; August 1993.
- Workshop on Comparison Geometry; Mathematical Sciences Research Institute, Berkeley, CA; October 1993.
- 4th Southern California Geometric Analysis Seminar, UC Irvine, CA; February 1995.
- GeomFest 1995; SUNY Stony Brook, NY; April 1995.
- Summer School on Geometry and Physics; Odense University, Denmark; July 1995.
- Mathematics and Finance Conference, Institute for Advanced Studies, Princeton, NJ; September 1997.
- SAS Conference on Data Mining Methods; Cary, North Carolina; September 1999.
- 6th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Boston, MA; August 2000.
- BioITWorld - the 1st Information Technology for Life Sciences Conference, Boston, MA; March 2002.
- Spam Conference, MIT, Cambridge, MA; January 2003.

- BioITWorld - the 2nd Information Technology for Life Sciences Conference, Boston, MA; March 2003.
- 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Washington, DC; August 2003.
- 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Seattle, WA; August 2004.
- BioITWorld - the 4th Information Technology for Life Sciences Conference, Boston, MA; April 2005.
- 11th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Chicago, IL; August 2005.
- 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Chicago, IL; August 2013.
- 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; New York, NY; August 2014.

Talks:

- International Student Conference, MAI, 1990.
- Department of Mathematics, Tufts University, 1996.
- ALPHATECH Inc, 2003.
- MIT - Lincoln Laboratory, 2010.
- NYC Machine Learning Meetup, 2015, 2016.
- MLConf 2016 Seattle.

Languages: English (fluent), Russian (native), French (reading), German (reading).

Independent software development:

- GNU Emacs - as a pre-tester, fixed numerous bugs and added some features, such as a package for automated maintenance of the add-on software etc. [1994-now]
- CLISP - an ANSI Common Lisp implementation; one of the maintainers, added some major functionality, such as PostgreSQL interface, multiple valued places, worked on standards compliance. [1997-now]
- SCWM - a Scheme-configurable X Window Manager; one of the maintainers, added some modules and an Emacs interface. [1997-1999]
- CLOCC - the Common Lisp Open Code Collection; one of the founders, maintain port (the cross-implementation portability package) and c11ib (the Common Lisp extension library). [1998-now]
- Implemented in **Lisp** a simple neural network and an *animals* player based on it. [1999-now]
- Developed an integrated software package facilitating foreign language learning. The package contained a **UNIX Shell** script, a **Perl** script and a **C** program. [1992-1995]
- Designed and implemented in **C++** a package integrating an interactive tool for John Horton Conway's evolution simulation game "life", and games Mastermind and Battleship. [1994-1995]
- Developed a transliteration/encoding program for Internet e-mail transmission (implemented in **C**). Originally designed for Cyrillic, it can be used for any non-ASCII character set. [1994-now]
- Designed and implemented in **C** a command line interface to a database created by a Lisp package. This involved heavy string parsing and optimization. [1995-1997]
- Designed a command line HTTP interface, and, based on it, a system for querying servers for stock quotes, keeping the history and graphing it using third party software via IPC. Implemented in **C**, **Perl**, **Java** (with a GUI) and **Lisp**. [1995-now]
- Designed and implemented in **Java** a GUI for interactive map creation via the WWW. [1997]
- Designed and implemented in **Java** an IMAP command-line client for mail retrieval automation. [1999-2000]