

# TOM FAWCETT

TOM.FAWCETT@GMAIL.COM

<http://tomfawcett.info>

## CAREER OBJECTIVES

Applied research and development in machine learning and data mining. Specific interests are in automated knowledge discovery, relational data mining, and the application of machine learning techniques to real-world problems.

## EDUCATION

UNIVERSITY OF MASSACHUSETTS AT AMHERST  
Amherst, MA — Ph.D. (Computer Science), 1993

RUTGERS UNIVERSITY  
New Brunswick, NJ — M.S. (Computer Science), 1988

WORCESTER POLYTECHNIC INSTITUTE  
Worcester, MA — B.S. (Computer Science), 1980

## EXPERIENCE

### PRINCIPAL DATA SCIENTIST

*Silicon Valley Data Science, Mountain View, CA* 2/2014–present  
Leading data scientist at start-up data science/engineering consulting firm. Responsible for setting project directions, coordinating with project managers, mentoring junior data scientists, preparing status reports and presentations, and recruiting/interviewing.

### WRITING SABBATICAL

*Mountain View, CA* 6/2011–2/2014  
Co-authored a book on data mining geared toward business applications (***Data Science for Business: What you need to know about data mining and data-analytic thinking***), in collaboration with Foster Provost of NYU/Stern. Published by O'Reilly Media August 2013. Book is currently top-selling O'Reilly data book, in use in over one hundred schools worldwide. See [data-science-for-biz.com](http://data-science-for-biz.com).

### SENIOR MACHINE LEARNING ARCHITECT

*Proofpoint, Inc.; Sunnyvale, CA* 7/2010–5/2011  
Senior Architect for an email management company, responsible for the application of machine learning techniques to email analysis and filtering. Applied social network analysis and probabilistic inference techniques to the problem of spam detection and filtering based on incoming header attributes. Resulting “Circle of Trust” prototype demonstrated ability to identify significant proportion of spam stream with high confidence, and to identify suspicious links worthy of further investigation.

### SABBATICAL

*Mountain View, CA* 11/2007–6/2010  
Research sabbatical devoted to investigating various topics of interest (swarm intelligence, cellular automata and complexity as they relate to machine learning). Continued responsibilities as action editor, reviewer, program chair, etc. Published several papers on these topics and edited a special issue of *Machine Learning Journal* on Swarm Intelligence for Data Mining.

### TECHNICAL PROJECT MANAGER AND TECHNICAL CONTRIBUTOR

*Institute for the Study of Learning and Expertise; Palo Alto, CA* 9/2005–11/2007  
Project manager for Stanford team (prime contractor) working on DARPA project related to transfer learning. Responsibilities included coordinating research groups at seven universities, establishing and monitoring project goals and procedures, coordinating with DARPA program management, giving presentations, and writing progress reports. Technical contributor for Stanford/ISLE project on DARPA transfer learning project.

### SENIOR RESEARCH SCIENTIST

*Hewlett-Packard Laboratories; Palo Alto, CA* 10/1999–1/2005  
Senior Research Scientist in Data Mining and Machine Learning Group, conducting applied research on problems of interest to Hewlett-Packard. Responsible for coordinating projects, setting technical directions, co-managing collaboration with business partners, writing reports and applying for patents.

Domains of the data mining and machine learning project included large-scale machine configuration analysis, text classification, credit card fraud detection, and targeted marketing analysis.

#### RESEARCH SCIENTIST

*NYNEX/Bell Atlantic Science and Technology; White Plains, NY*

*8/1993–9/1999*

Co-leader of the machine learning project, which applied machine learning and data mining techniques to telecommunications problems. Solely responsible for design and implementation of project software (in Perl and C). Jointly responsible for setting research directions, writing research proposals and reports, authoring papers and preparing presentations. Co-administrator of research grants with three universities.

Domains of the machine learning project included detection of cellular phone cloning fraud, diagnosis of local-loop phone troubles, prediction of T1 digital switch failures, and statistical analysis of NYNEX trouble rate. This work led to numerous papers, a patent, and a NYNEX President's Award.

#### TECHNICAL STAFF MEMBER

*GTE Laboratories Inc.; Waltham, MA*

*4/1983–9/1985*

*During grad school: Summers of 1987–1991*

Conducted applied research and development in machine learning. Developed learning agents for telecommunications network management, as part of the department's Integrated Learning System. Prepared technical reports, research summaries and presentations. Developed and implemented a telecomm network simulator to support agent testing and development.

#### TECHNICAL STAFF MEMBER

*The MITRE Corporation; Bedford, MA*

*7/1980–3/1983*

Developed and maintained a large knowledge-based expert system for mission planning, implemented in Lisp. Designed and implemented forward-chaining rule interpreter, frame system and constraint interpreter. Pursued research in planning and knowledge representation. Assisted with demonstrations and presentations. Wrote substantial portions of quarterly and yearly reports for sponsor.

#### TECHNOLOGIES

Perl, Python, SQL, LISP, Prolog, FORTRAN  
Hadoop, Spark  
Python, Pandas, Scikit-learn, PySpark  
Deep Learning (Keras, Tensorflow)  
Familiar: Java, C, Ruby, R

#### AWARDS

President's Award, *Report Rate Study and Diagnosis*, NYNEX Science and Technology. 1995.

Best Paper Award (Fundamental Research), "Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions." *Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, 1997.

SCOPUS 1000 Award (Highly cited paper), "An introduction to ROC analysis." *Pattern Recognition Letters*, 2012.

## PROFESSIONAL SERVICE

### Program Chair

Twentieth International Conference on Machine Learning (ICML-2003)  
Second International Conference on Email and Anti-Spam (CEAS-2005)

### Editor

Machine Learning Journal Action Editor (Jan 2004 – Dec 2010).  
Guest editor of Machine Learning Special Issue on Data Mining Lessons Learned (2004)  
Guest editor of Machine Learning Special Issue on Swarm Intelligence for Knowledge Discovery in Data (2009).

**Editorial Board**, Machine Learning Journal (Jan 2003 – Jan 2004; Jan 2011 – present)

**Area Chair** (Applications of machine learning), ICML-2000, KDD-2007.

**Tutorials Chair**, Int'l Conf on Knowledge Discovery and Data Mining (KDD-2001).

### Workshop Organizer

ICML-02 Workshop: Data Mining Lessons Learned. (Co-Chair)  
ICML/AAAI-98 Workshop: AI Approaches to Time-series Problems. (Co-Chair)  
AAAI-97 Workshop: AI Approaches to Fraud Detection and Risk Management. (Chair)  
ICML-94 Workshop: Constructive Induction and Change of Representation. (Chair)

### Program Committee Member

ICML-05 (Int'l Machine Learning Conference), ICDM-05 (IEEE Conference on Data Mining), ECAI-04 (European Conference on AI) Workshop on ROC Analysis in AI, CEAS-04 (First Conference on Email and Anti-Spam), KDD-02 (Int'l Conference on Knowledge Discovery and Data Mining), KDD-01, KDD-99, KDD-98 (also served on Best Paper Awards committee) SCDM-02 (SIAM Second International Conference on Data Mining) RDBM-01 (Int'l Workshop on Rule-based Data Mining) ICDM-01 (IEEE Conference on Data Mining), DELOS/NSF Workshop on Personalization and Recommender Systems in Digital Libraries.  
Various AAAI, ECAI and ICML conferences.  
IEEE Conference on Tools with Artificial Intelligence (1993, 1996).

### Reviewer

*Conferences*: AAAI, ECAI, SigKDD  
*Journals*: Machine Learning, Journal of Artificial Intelligence Research, Computational Intelligence, Data Mining and Knowledge Discovery  
*Proposals*: National Science Foundation

**Advisory Committee**, ICML-2004.

## PATENTS

Fawcett and Provost, "Automatic Design of Fraud Detection Systems."  
Patent 5,790,645. Granted 4 August 1998.

Forman, Fawcett and Suermondt, "Method and System for Measuring the Quality of a Hierarchy."  
Patent 7,043,468. Granted 9 May 2006.

Goin, Campbell, Stinger, Fawcett, Steele, Mishra and Suermondt, "Method and system for clustering computers into peer groups and comparing individual computers to their peers."  
Patent 7,203,864. Granted 10 April 2007.

Goin, Campbell, Stinger, Fawcett, Steele, Mishra and Suermondt, "Method and system for comparing individual computers to cluster representations of their peers."  
Patent 7,380,177. Granted 27 May 2008.

Fawcett, Forman and Suermondt, "Identifying Exceptional Managed Systems."  
Patent 7640217. Granted 31 Dec 2009.

## PUBLICATIONS

### BOOKS

*Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. F. Provost and T. Fawcett. O'Reilly Media, Aug 2013. <http://www.data-science-for-biz.com>

*Proceedings of the Twentieth International Conference on Machine Learning*. T. Fawcett and N. Mishra (eds.) AAAI Press, 2003.

### JOURNAL ARTICLES

F. Provost and T. Fawcett, "Data Science and its Relationship to Big Data and Data-Driven Decision Making". *Big Data* v.1, n.1, 2012.

D. Martens, B. Baesens and T. Fawcett, "Editorial survey: Swarm Intelligence for Data Mining". *Machine Learning*, Vol. 82, No. 1, Springer Netherlands, Jan. 2011

T. Fawcett, "PRIE: A system for generating rulelists to maximize ROC performance". *Data Mining and Knowledge Discovery*, Vol. 17, No. 2, Oct. 2008.

T. Fawcett and Alexandru Niculescu-Mizil. "PAV and the ROC convex hull", *Machine Learning*, Vol. 68, No. 1, July 2007.

T. Fawcett, "ROC graphs with Example-Specific Costs." *Pattern Recognition Letters*, Vol. 27, Issue 8, June 2006.

T. Fawcett, "An introduction to ROC analysis." *Pattern Recognition Letters*, Vol. 27, Issue 8, June 2006. **Winner of Scopus1000 award in recognition of having received over 1,000 citations** (1,126 citations as of Jan. 2012).

T. Fawcett and P. Flach, "A Response to Webb and Ting's *On the Application of ROC Analysis to Predict Classification Performance under Varying Class Distributions*". *Machine Learning*, Vol. 58, No. 1, 2005.

N. Lavrač, H. Motoda, T. Fawcett, R. Holte, P. Langley and P. Adriaans. "Lessons Learned from Data Mining Applications and Collaborative Problem Solving". *Machine Learning Special Issue on Data Mining Lessons Learned*, Vol. 57, No. 1-2, Oct. 2004.

F. Provost and T. Fawcett, "Robust Classification for Imprecise Environments." *Machine Learning*, Vol. 42, No. 3, March 2001.

T. Fawcett and F. Provost, "Adaptive Fraud Detection." *Data Mining and Knowledge Discovery*, Vol. 1, No. 3, Nov. 1997.

T. Fawcett, "Knowledge-based Feature Discovery for Evaluation Functions." *Computational Intelligence*, Vol. 12, No. 1, Feb. 1996.

J. Vittal, B. Silver, W. Frawley, G. Iba, T. Fawcett, S. Dusseault and J. Doleac, "Intelligent and Cooperative Information Systems Meet Machine Learning." *The International Journal of Intelligent and Cooperative Information Systems*, Vol. 1, No. 2, 1992.

### REFEREED CONFERENCE PAPERS

T. Fawcett, "Using Rule Sets to Maximize ROC Performance". *IEEE International Conference on Data Mining (ICDM-01)*, Nov. 2001.

T. Fawcett and F. Provost, "Activity Monitoring: Noticing Interesting Changes in Behavior." *Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99)*, Aug. 1999.

F. Provost, T. Fawcett and R. Kohavi, "The Case Against Accuracy Estimation for Comparing Induction Algorithms." *Fifteenth International Conference on Machine Learning (ICML-98)*, July 1998.

F. Provost and T. Fawcett, "Robust Classification Systems for Imprecise Environments." *Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, July 1998.

F. Provost and T. Fawcett, "Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions." *Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, Aug. 1997.

**Awarded prize for Best Fundamental Research Paper.**

T. Fawcett and F. Provost, "Combining Data Mining and Machine Learning for Effective User Profiling." *Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, Aug. 1996.

T. Fawcett and P. Utgoff, "Automatic Feature Generation for Problem Solving Systems." *Ninth International Conference on Machine Learning*, Aug. 1992.

J. Callan, T. Fawcett and E. Rissland, "CABOT: An adaptive approach to case-based search." *Twelfth International Joint Conference on Artificial Intelligence*, Aug. 1991.

T. Fawcett and P. Utgoff. "A Hybrid Method for Feature Generation." Proceedings of the Eighth International Workshop on Machine Learning, July 1991.

#### BOOK CHAPTERS AND OTHER PUBLICATIONS

T. Fawcett, "Data mining with cellular automata". *SigKDD Explorations*, Vol. 10, No. 1, June 2008.

N. Lavrač, H. Motoda and T. Fawcett, "Editorial: Data mining lessons learned". *Machine Learning Special Issue on Data Mining Lessons Learned*, Vol. 57, No. 1-2, Oct. 2004.

T. Fawcett, "In vivo Spam Filtering: A Challenge Problem for Data Mining". *SigKDD Explorations*, Vol.5, No.2. Dec. 2003.

T. Fawcett. "ROC graphs: Notes and practical considerations for researchers". HP Labs Tech Report HPL-2003-4.

T. Fawcett and F. Provost, "Fraud Detection." *Handbook of Data Mining*, edited by Jan Zytkow and Willi Klösgen. Oxford University Press. 2002.

T. Fawcett, "Case study: Adaptive fraud detection." *Handbook of Data Mining*, edited by Jan Zytkow and Willi Klösgen. Oxford University Press. 2002.

A. Danyluk, T. Fawcett and F. Provost, "AI Approaches to Time-series Problems." Workshop report for *AI Magazine*. 2000.

T. Fawcett, I. Haimowitz, S. Stolfo and F. Provost, "AI Approaches to Fraud Detection and Risk Management." Extended workshop report in *AI Magazine*, 1998.

T. Fawcett, "Feature Discovery for Problem Solving Systems." Doctoral dissertation, Department of Computer Science, University of Massachusetts. 1993. Technical Report TR-93-49.

B. Silver, J. Vittal, W. Frawley, G. Iba, T. Fawcett, S. Dusseault and J. Doleac, "A Framework for Integrating Heterogeneous Learning Experts." In *Second Generation Expert Systems*. Edited by Jean-Marc David, Jean-Paul Krivine, and Reid Simmons. Springer Verlag, 1993.

J. Vittal, B. Silver, W. Frawley, G. Iba, T. Fawcett, S. Dusseault and J. Doleac, "A Framework for Cooperative Adaptable Information Systems." In *The Next Generation of Information Systems - From Data To Knowledge*. Edited by Mike P. Papazoglou and John Zeleznikow. Springer Verlag, 1992.

J. Callan, T. Fawcett and E. Rissland, "Adaptive case-based reasoning." *The 1991 DARPA Workshop on Case-Based Reasoning*, Washington, D.C., May 1991.

T. Fawcett, "Learning from plausible explanations." *Sixth International Workshop on Machine Learning*, Ithaca, NY, Aug. 1990.