Foster Provost
fprovost@stern.nyu.edu

RESEARCH INTERESTS

Machine Learning & AI
Data Science & Business Analytics
Causal Prediction

Human-AI Integration
Advertising and Advertising Technology
(Social) Network Data

EMPLOYMENT HISTORY

Distinguished Scientist, Compass, 2019-present

Ira Rennert Professor of Entrepreneurship, Stern School of Business, NYU, 2020-present
Professor of Data Science, New York University, 2015-present
Professor of Information Systems, Stern School of Business, NYU, 2008-present

Director, Data Analytics & AI, Fubon Center, Stern School, NYU, 2018-present
Director (interim), Center for Data Science, NYU, 2015-2016

Associate Professor of Information Systems, Stern School of Business, New York University, 2001-2008
Assistant Professor of Information Systems, Stern School of Business, New York University, 1999-2001

Research Scientist, NYNEX Science and Technology, Inc. (now Verizon), 1994-1999

HONORS & AWARDS

2020 ACM SIGKDD Test of Time Award – *for the paper "Get Another Label…", which has had a "lasting impact on the research community and continues to be cited as the foundation for new branches of research"*

Best Student Paper Award Runner-up, WISE 2020 (student: Carlos Fernandez)

2019 Company of the Year - CFO Info Tech – for Detectica (Founder with Ipeirotis and Attenberg)

Finalist, 2019 INFORMS Case Competition (*Data-Driven Investment Strategies for Peer-to-Peer Lending*)

Finalist, 2018 Algorithm for Change AI/ML Competition *(Helping low income, underrepresented minority and first generation students get to and through college)*

European Research Paper of the Year 2017 (awarded by AIS & CIONET – for a paper in *MISQ*)

Best Paper Award, *Information Systems Research* (2016)

INFORMS President's Pick, for paper published in *Information Systems Research* (Fall 2015)

2014 NYU/Stern MSBA Best Teacher Award

3 of the Top-5 Most Downloaded Papers (including #1) in the journal *Big Data* (reported Nov. 2015)

Nominated for 2014 NYU/Stern Professor of the Year by the MBA student body (one of 31 nominees)

Nominated for 2013 NYU/Stern Professor of the Year by the MBA student body (one of 23 nominees)

Best Paper Award, ACM SIGKDD 2012, Industry Track

People's Choice Award, Runner-up, 2012 Future of Advertising Conf. on Empirical Generalizations in Advertising

Wharton Customer Analytics Initiative Awards, 2011, 2012, 2014 (one of several)

Marketing Science Institute Award, 2009 (one of several)

The INFORMS Design Science Award, December 2009

Best Paper Award Runner-up, ACM SIGKDD 2008

Elected as one of ten founding board members of the Intl. Machine Learning Society, Fall 2001 (Reelected 2006)

Second Place in KDDCUP-2003; Third Place in KDDCUP-2003

IBM Faculty Awards, 2000 & 2001

Best Paper Award, Third International Conference on Knowledge Discovery and Data Mining (KDD-97)

President's Award, NYNEX Science and Technology (now Verizon), 1995 (for a predictive analytics model that prioritized network rehabilitation for maximal impact)

Graduated First in Class, Duquesne University, 1986. Summa Cum Laude (4.0 GPA), and awarded the Certificates of General Excellence, Excellence in Mathematics, and Excellence in Physics

EDUCATION

Ph.D.   Computer Science, University of Pittsburgh, 1993
M.S.   Computer Science, University of Pittsburgh, 1988
B.S.   Physics and Mathematics, Duquesne University, 1986

ADVISORY POSITIONS

Scientific Advisory Board, ISI Foundation (awards the Lagrange Prize, among other things), 2015-2020

NYU Venture Fund Investment Review Board, 2013-

Academic Advisory Committee, NYU Skirball Center for the Performing Arts, 2015-2017

Wharton Customer Analytics Research Council, 2011-2015

Academic Advisor, NYC Media Lab, 2011-2014

Chief Scientist (Consulting), Coriolis Ventures, 2008-2011

DARPA Workshop on Network Science, at the intersection of Economics and Machine Learning, October 2005.

U.S. Treasury Department (future efforts towards Resilient Finanical Information Systems), March 2005.

National Academies' Computer Science and Telecommunications Board planning workshop on information fusion for counter-terrorism, June 2002.

White House Office of Science and Technology Policy (data mining for counter-terrorism), February 2002

COMPANIES coFounded in one sense or another

Dstillery (data-science-based advertising; formerly Media6Degrees), 2008-
  2014 Forbes' America's Most Promising Companies (#47)

Integral Ad Science (data science/analytics; formerly AdSafe Media), 2009-(2018)
  IPO 2021
  Acquired by Vista Equity Partners for a reported $850M in July 2018
  2015 Forbes' America's Most Promising Companies (#15)
  Crain's New York Business 6ᵗʰ fastest growing company in NY (2013)
  2012 Product of the Year - Business Intelligence Group Awards (Professional Services)

Everyscreen Media (data-driven mobile advertising), 2011-2013
  Acquired by Media6Degrees, May 2013

Predicube (data-driven advertising in Europe), 2014-2018
  Acquired by Var (Belgium), August 2018
  2016 Annual Master of Media Award - First place for Best Innovation (United Media Agencies, Belgium)

Detectica (controlling financial malfeasance + helping firms with their AI/Data Science strategy), 2015-2019
  Acquired by Compass, October 2019 (Compass IPO'ed April 2021)
  2019 Company of the Year - CFO Info Tech
  Top-10 AI Solution Providers 2019 - CFO Info Tech
  Awarded Most Innovative Trader Surveillance Solutions Developer 2019 - New York by Acquisition INTL
  Awarded New York's Most Progressive AI and Data Science Firm 2020, by Acquisition INTL
  2017 Graduate FinTech Innovation Lab (one of seven startups selected)

UNIVERSITY/SCHOOL/DEPARTMENTAL SERVICE

University: Director (interim) of NYU's Center for Data Science (2015-2016); NYU Committee on Intellectual Property Policy (2013-2018); Steering Committee for NYU Moore Sloan Data Science Environment Initiative--$37.8M program with Berkeley and U.Wash. (2013-2016); NYU Venture Fund Investment Review Board (2013-); Skirball Center Academic Advisory Committee (2015-); NYU MSDS Curriculum Committee (2014-2016); University Working Group for 2014 Middle States Self-Study (2012-2014); University Working Group on Data Science Center and Data Science degree programs (2012-2013);

School/Department: Director: Fubon Center, Data Analytics and AI (2018-); Chair of Faculty Recruiting, IOMS/IS (2018-); Creative Destruction Lab Scientist; Faculty Liaison/Advisor to the Stern Business Analytics

Clubs (2017-2018), co-designed and organized analytics competitions, panels, and other events. Co-designed & Hosted MSBA Business Analytics Symposium, "Competing in the Connected Economy: Applications & Implications" (Fall 2016); Design & Host Stern Biz Analytics Club event, "Whiskey Analytics (with tasting)" (each year); Stern MSBA curriculum committee (2015-2016); IOMS/IS curriculum committee (2014-); Co-organized the IOMS Dept. Research Fest, Fall 2014; IS Group faculty recruiting committee (2012-2013); Statistics Group faculty recruiting committee (2010-2011); Deans' Faculty Advisory Committee (2007-2010); IS Group faculty recruiting committee (2000-2010); Stern Ph.D. Program Oversight Committee (2008, 2004); Cochair of faculty recruiting committee for IS Group (2007-2008); Chair of faculty recruiting for IS Group (2003-2005); Ph.D. recruiting committee for IS group (2007); Led the design of proposal for school-wide incentive policy for faculty procurement of external funding (2007-2009); MBA Portfolio Committee (2004-2005); U.G. Dean's Faculty Advisory Committee (2004-2005); *and others going back further*.

SELECTED RECENT TEACHING & COURSE DEVELOPMENT

MSQM Class: Predictive Analytics, Developed (with P. Ipeirotis) 2020.
MSBA Class: Introduction to Business Analytics, Developed & Taught 2017-2021
MBA Class: Data Science for Biz Analytics (Technical), Developed & Taught Spring 2017-2021
MBA Class: Data Mining for Biz Analytics (Managerial): Developed & Taught Fall 2016-2017
Ugrad Class: Data Mining for Business Analytics: Developed & Taught Spring 2007-2011; S2018
TRIUM Exec Education Class: Data Analytics, Developed & Taught 2015-2016
MSDS/MBA Class: Intro to Data Science/Data Mining for Biz Analytics, Developed & Taught Fall 2013, 2015
MSBA Class: Data Science for Business Analytics: Developed & Taught 2013, 2014, 2016
MBA Class: Practical Data Science: Developed & Taught (with Josh Attenberg) Fall 2012.
MBA Class: Data Mining for Business Analytics: Developed & Taught Fall 2005, Spring 2007-2013
MBA/PhD Class: Networks, Crowds & Markets; Developed and Taught (with lots of help) S2011
Ph.D. Seminar: Data Science Research. Developed & Taught: S2012, S2013, S2014, S2017
Ph.D. Seminar: Technical Information Systems Research. Developed & Taught: S2000, etc.
MBA Core Class: Information Technology and Electronic Commerce. Taught: S2000, S&F2001, S&F2002
MBA Class: Information and Internet Technologies. Developed (w/ V.Vassalos): S2000. Taught: S2000, F2000

GRANTS

- PI: Marketing Science Institute Award ($6500) "On-line brand advertising using social networks induced from user-generated content," 2009.
- Co-PI: IBM OCR Award ($200,000) "Data and Risk." 2008-2010.
- PI: DOD/AFRL Grant ($186,925) Evidence Assessment, Grouping, Linking, and Evaluation Program, 2004-5.
- PI: NYSIA Grant ($35,000), "Algorithms and software for suspicion scoring," 2004-2005.
- PI: NSF Grant IIS-0329135 ($67,087), "Generation & Evaluation of Confidence Bounds for ROC Curves," 2003-2004.
- PI: DARPA Grant ($825,000), Evidence Extraction and Link Discover Program, 2001-2003.
- PI: IBM Faculty Awards ($80,000), "Extending the Reach of Data Mining," 2000, 2001.
- Co-PI: NSF Grant NSF-IRI-9412549 ($240,101), "Using Parallelism to Scale Up Machine Learning to Large Data-Analysis Problems" 1994.

PATENTS

- System and Method for Estimating Audience Interest (#10,599,981 B2)
- Methods, Systems, and Media for Identifying Errors in Predictive Models using Annotators (#9,311,599 B1) (#10,846,600)
- Privacy-sensitive Methods, Systems, and Media for Geo-Social Targeting (#10,034,135)
- *Methods for enhancing privacy by inhibiting unwanted inferences*: System, method and computer-accessible medium for predicting private traits from digital indicators (#9,767,468)
- Apparatus, method and computer-accessible medium for explaining classifications of documents (#9,836,455)
- Methods, systems, and media for providing hybrid learning approaches (pending)
- Methods, systems, and media for determining location information from real-time bid requests (#9,014,717) (#9,179,264)
- Methods, systems, and media for applying scores and ratings to web pages, web sites, and content for safe and effective online advertising (#8,732,017)
- System, process and software arrangement for assisting with a knowledge discovery process (#7,167,846)
- Automated design of fraud detection systems (#5,790,645)

FACULTY FELLOWSHIPS

- Andre Meyer Faculty Fellow (2017-2020)
- Andre Meyer Faculty Fellow (2014-2017)
- NEC Faculty Fellow (2011-2014)
- NEC Faculty Fellow (2008-2011)
- Paduano Fellow in Business Ethics (2007-2008)

Ph.D. STUDENTS & POSTDOCS

Maytal Saar-Tsechansky, Associate Professor, University of Texas at Austin
  Stern School, NYU, Ph.D. 2002
  Outstanding Winner (first place), eBRC/SAP Doctoral Award, 2001

Claudia Perlich, Senior Data Scientist, TwoSigma (formerly at IBM Research and Chief Scientist, Dstillery)
  Stern School, NYU, Ph.D. 2005
  Winner (one of four), eBRC/SAP Doctoral Award, 2002

Shawndra Hill, Principal Scientist and Manager at Facebook and Senior Lecturer at Columbia Business School
  Stern School, NYU, Ph.D. 2007
  Winner of the Herman E. Kroos Award for the best dissertation at NYU's Stern School, 2007
  Finalist, George B. Dantzig Dissertation Award (INFORMS)

Josh Attenberg, CTO Resonance (formerly Director, Data Science at Etsy)
  NYU/Poly (Tandon) Computer Science Ph.D., 2014

Xiaohan Zhang, Machine Learning Engineer, Sony (formerly Director, Data Science, Integral Ad Science)
  NYU/Stern Ph.D., 2014

Jessica Clark, Assistant Professor, University of Maryland
  Stern School, NYU, Ph.D. 2017

Robert Moakler, Research Scientist, Facebook
  Stern School, NYU, Ph.D. 2017

Carlos Fernández-Loría, Assistant Professor, HKUST (Fall 2021)
  Winner of the Herman E. Kroos Award for the best dissertation at NYU's Stern School, 2021
  Stern School, NYU, Ph.D. 2021

Enric Junqué de Fortuny. Assistant Professor, NYU Shanghai
  Univ. Antwerp Ph.D. 2014 (Co-advised with Prof. D. Martens of Univ. Antwerp)

Gary Weiss, Associate Professor, Fordham University, Dept. of Computer and Information Sciences
  Ph.D. Rutgers University 2003 (Co-advised with Prof. H. Hirsh of Rutgers University)

Sofus Macskassy, VP Data Science, HackerRank
  Formerly Director, Fetch Labs, Fetch Technologies & Head of Data Analytics, Branch Metrics
  Associate Research Scientist, Stern School. 2004-2005
  Ph.D. Rutgers University 2003

Victor Shengli Sheng, Associate Professor, Univ. Central Arkansas
  Associate Research Scientist, Stern School. 2007-2009.
  Ph.D. Univ. W. Ontario, 2007

PROFESSIONAL SERVICE

University External Review Committees

President's Advisory Board, Carnegie Mellon University, *for the review of CMU's Heinz College* (2019)

External Review Committee, Stevens Institute, *for the reviews of MIS and Business Analytics* (2019)

Editor-in-Chief:

Machine Learning journal (January 2004 – June 2010)

Founding Board Member (elected):

International Machine Learning Society (2001-2010)

Program Chair/General Chair/Main Organizer:

IEEE DSAA 2018 - The 5th IEEE International Conference on Data Science and Advanced Analytics. (General Chairs –Francesco Bonchi and Foster Provost)

The 1st, 2nd, 3rd, 4th, 5th and 6th Workshops on Information in Networks (WIN), Sept 2009-2013, 2015. (Social Networks Summit among researchers from Computer Science, Economics, Management, Physics, Sociology, and other disciplines.) (co-organizer/co-chair with Sinan Aral & Arun Sundararajan)

The 1st & 2nd Human Computation Workshops (HCOMP-2009, -2010) (First workshops on "human computation" – crowdsourcing, micro-outsourcing, etc. for solving large or difficult problems; co-located with SIGKDD 2009 & SIGKDD 2010) (co-founder/co-organizer/co-chair with several others)

The 1st Workshop on Social Media Analytics (SOMA-2010) (co-organizer/co-chair with Prem Melville & Jure Leskovec)

Fourth Symposium on Statistical Challenges in Electronic Commerce Research (SCECR 2008) (co-organizer/co-chair with A. Ghose, and A. Sundararajan)

Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001) (Program Chair – with R. Srikant)

Editorial Boards:

Data Mining and Knowledge Discovery journal [2007-present]
Machine Learning journal [199?-present] (Action Editor, 2001-2003)(Editor-in-Chief, 2004-2010)
Journal of Machine Learning Research (JMLR) [2000-2010]
Journal of Artificial Intelligence Research (JAIR) [2002-2005]

Guest Editor:

Data Mining and Knowledge Discovery 5 (1/2) (January/March 2001). Special issue on "Applications of Data Mining to Electronic Commerce" (with Ron Kohavi).

Machine Learning 30 (2/3) 1998. Special issue on "Applications of Machine Learning and the Knowledge Discovery Process" (with Ron Kohavi).

PUBLICATIONS

PAPERS SUBMITTED/UNDER REVISION/IN PREPARATION

1. "Who's Watching TV?" J. Clark, J.-F. Paiement, F. Provost. *Under revision*.
2. "Observational vs Experimental Data When Making Automated Decisions Using Machine Learning" C. Fernandez & F. Provost. *Under revision*.
3. "Methods for Individual Treatment Assignment: An Application and Comparison for Playlist Generation." C. Fernandez, F. Provost , J. Anderton, B. Carterette, & P. Chandar. *Under review*.

MUSIC

4. Hurvich, C. and F. Provost. *Mean Reversion*. (Original Progressive Rock) Album released October 2018 by Kloster Art Music.

5. Hurvich, C. and F. Provost. *Blessing of My Birth*. (Original Progressive Rock) Single released September 2018 by Kloster Art Music.

6. F. Provost and A. Selvin. "Never Going to Work Again." Single released August 2018 by Kloster Art Music.

BOOKS

7. Provost, F. and T. Fawcett. ***Data Science for Business: What you need to know about data mining and data analytic thinking***. O'Reilly Media, 2013. **As of early 2017, adopted by >150 universities in at least 36 countries,** including Cambridge, Carnegie Mellon, Duke, Georgia Tech, Harvard, INSEAD, MIT, Stanford, U. Minnesota, U. Texas Austin, Wharton, and others. (see http://data-science-for-biz.com for a partial list). **One of *Fortune* Magazine's five "must read books for MBAs", December 2014.**

8. Provost, F. and R. Srikant (Eds.). *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, 2001.

9. Kohavi, R. and F. Provost (Eds.). *Applications of Data Mining to Electronic Commerce*. Kluwer Academic Publishers, 2001.

INVITED WHITE PAPER

10. "Big Data, Data Science, and Civil Rights." Barocas, Bradley, Honavar, and Provost (**2017**). *Invited paper for the Computing Community Consortium of the Computing Research Association (CRA)*. http://cra.org/ccc/resources/ccc-led-whitepapers/

BLOG POSTS

11. See https://medium.com/@ai_editors

JOURNAL ARTICLES

12. "Explaining AI Decisions: The Counterfactual Approach." C. Fernandez, F. Provost & X. Han. *Accepted to MISQ, 2021*. (Version available on arXiv.)

13. "Causal Decision Making and Causal Effect Estimation Are Not the Same... and Why It Matters." C. Fernandez & F. Provost. *INFORMS Journal on Data Science*, in press. (Version available on arXiv.)

14. "Causal Classification: Treatment effect estimation vs. outcome estimation." C. Fernandez & F. Provost. *Accepted (minor revisions) to the Journal of Machine Learning Research*. (Version available on SSRN.)

15. "Classification over bipartite graphs through projection." M. Stankova, S. Praet, D. Martens, and F. Provost. *Machine Learning* (2020).

16. "Instance-level explanation algorithms SEDC, LIME, SHAP for behavioral and textual data: a counterfactual-oriented comparison." Y. Ramon, D. Martens, F. Provost, & T. Evgeniou. *Advances in Data Analysis and Classification* 4/2020.

17. "A Benchmarking Study of Classification Techniques for Behavioral Data." S. De Cnudde, D. Martens, T. Evgeniou & F. Provost. *Int J Data Sci Anal* 9(2)**,** 131–173 (2020).

18. "Deep Learning on Big, Sparse, Behavioral Data." S. De Cnudde, Y. Ramon, D. Martens & F. Provost (2019). *Big Data* 7(4): 286-307.

19. "Unsupervised Dimensionality Reduction vs. Supervised Regularization for Classification from Sparse Data." J. Clark & F. Provost (2019). *Data Mining and Knowledge Discovery* 33(4):871–916.

20. "Data-Driven Investment Strategies for Peer-to-Peer Lending." M. Cohen, C.D. Guetta, K. Jiao, F. Provost (2018). *Big Data* 6(3), 191-213. (When last checked, **the journal's 4th most downloaded paper.**)

21. "Wallenius Bayes." E. Junqué de Fortuny, D. Martens, and F. Provost (2018). *Machine Learning* 107 (6): 1013–1037.

22. "In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions." E. Ascarza, S. Neslin, O. Netzer, Z. Anderson, P. Fader, S. Gupta, B. Hardie, A. Lemmens, B. Libai, D. Neal, F. Provost, R. Schrift (2017). *Customer Needs and Solutions*. Also: Marketing Science Institute Working Paper Series 2017 Report No. 17-118.

23. "Enhancing Transparency and Control When Drawing Data-Driven Inferences About Individuals" D. Chen, S. Fraiberger, R. Moakler & F. Provost (2017). *Big Data* 5(3): 197-212.

24. "Cost-Effective Quality Assurance in Crowd Labeling." J. Wang, P. Ipeirotis, and F. Provost (2017). *Information Systems Research* 28(1), 137-158. http://pubsonline.informs.org/doi/10.1287/isre.2016.0661

25. "What Managers Need to Know About Big Data." Foster Provost & Jim Euchner. *Research Technology Management* 60(3), 11-17, 2017.

26. "Mining Massive Fine-Grained Behavior Data to Improve Predictive Analytics." Martens, D., F. Provost, J. Clark, E. Junqué de Fortuny. *MIS Quarterly, Vol. 40 No. 4, pp. 869-888/December 2016.* **European Research Paper of the Year 2017** *(organized by AIS and CIONET)*

27. "Finding Mobile Consumers with a Privacy-Friendly Geo-Similarity Network." Provost, F., D. Martens, and A. Murray. *Information Systems Research* 26(2), 243-265, 2015. **Best Paper Award – ISR 2016**. **Also**, **INFORMS President's Pick** (Fall 2015).

28. "Evaluating and Optimizing Online Advertising: Forget the click, but there are good proxies." B. Dalessandro, R. Hook, C. Perlich, F. Provost. (Preliminary version won **People's Choice Award** Runner-up at the 2012 Empirical Generalizations in Advertising Conference.*) Big Data* 3(2), 90-102, 2015.

29. "Beat the Machine: Challenging Humans to Find a Predictive Model's 'Unknown Unknowns.'" Attenberg, J., P. Ipeirotis, F. Provost. *ACM Journal of Data and Information Quality* 6(1) 2015.

30. "A Data Scientist's Guide to Startups." F. Provost, G. Webb, R. Bekkerman, O. Etzioni, U. Fayyad, C. Perlich. *Big Data* **2**(3):117-128, September 2014.

31. "Machine learning for targeted display advertising: Transfer learning in action." C. Perlich · B. Dalessandro · O. Stitelman · T. Raeder · F. Provost. *Machine Learning* 95(1):103-127, 2014.

32. "Explaining Data-Driven Document Classifications." Martens, D. and F. Provost. *MIS Quarterly* 38(1), 73-99, 2014.

33. "Repeated Labeling Using Multiple Noisy Labelers." Ipeirotis, P., F. Provost, V. Sheng, J. Wang. *Data Mining and Knowledge Discovery* 28(2), 402-441, 2014.

34. "Predictive Modeling with Big Data: Is Bigger Really Better?" E. Junqué de Fortuny, D. Martens, and F. Provost. *Big Data* 1(4), 215-226, 2013 (**the journal's 6th most downloaded paper**).

35. "Information in Digital, Economic and Social Networks." Sundararajan, A., F. Provost, G. Oestreicher-Singer, and S. Aral. *Information Systems Research* 24(4), 883-905, 2013.

36. "Data Science and its Relation to Big Data and Data-driven Decision Making." F. Provost and T. Fawcett. *Big Data* 1(1), 51-59, 2013 (**the journal's #1 most downloaded paper**).

37. "Inactive Learning? Difficulties Employing Active Learning in Practice." Attenberg, J. and F. Provost. *SIGKDD Explorations* 12(2), 2010.

38. Saar-Tsechansky, M., P. Melville, and F. Provost. "Active Feature-value Acquisition." *Management Science* 55(4): 664-684, 2009.

39. Saar-Tsechansky, M. and F. Provost. "Handling Missing Features when Applying Classification Models." *Journal of Machine Learning Research* 8(Jul):1623-1657, 2007.

40. Macskassy, S. and F. Provost. "Classification in Networked Data: A toolkit and a univariate case study." *Journal of Machine Learning Research* 8(May):935-983, 2007.

41. Saar-Tsechansky, M. and F. Provost. "Decision-centric Active Learning of Binary-Outcome Models" *Information Systems Research* 18(1), March 2007, pp. 4-22.

42. Hill, S., F. Provost, and C. Volinsky. "Network-based marketing: Identifying likely adopters via consumer networks." *Statistical Science* 21(2), 256-276, 2006.

43. Perlich, C. and F. Provost. "Distribution-based Aggregation for Relational Learning from Identifier Attributes." *Machine Learning* 62 (1/2) 65-105, 2006.

44. Bernstein, A., F. Provost and S. Hill. "Intelligent Assistance for the Data Mining Process: An Ontology-based Approach." *IEEE Transactions on Knowledge and Data Engineering* 17(4), pp. 503-518. April 2005.

45. Saar-Tsechansky, M. and F. Provost. "Active Sampling for Class Probability Estimation and Ranking." *Machine Learning* 54(2) 2004, 153-178.

46. Hill, S. and F. Provost. "The Myth of the Double-Blind Review? Author Identification Using Only Citations." *SIGKDD Explorations* 5(2) 2003, 114-119.

47. Perlich, C., F. Provost, and S. Macskassy. "Predicting citation rates for physics papers: Constructing features for an ordered probit model." *SIGKDD Explorations* 5(2) 2003, 89-90.

48. Weiss, G. and F. Provost. "Learning when Training Data are Costly: The effect of class distribution on tree induction." *Journal of Artificial Intelligence Research* 19 (2003) 315-354.

49. Perlich, C., F. Provost, and J. Simonoff. "Tree Induction vs. Logistic Regression: A Learning-curve Analysis." *Journal of Machine Learning Research* 4 (2003) 211-255.

50. Provost, F. and P. Domingos. "Tree Induction for Probability-based Ranking." *Machine Learning* 52:3, 199-215, 2003.

51. Provost, F. and T. Fawcett. "Robust Classification for Imprecise Environments." *Machine Learning* 42, 203-231, 2001.

52. Dhar, V., D. Chou, and F. Provost, "Discovering Interesting Patterns for Investment Decision Making with GLOWER -- A Genetic Learner Overlaid With Entropy Reduction," *Data Mining and Knowledge Discovery* 4(4): 251-280, 2000.

53. Provost, F. and V. Kolluri, "A Survey of Methods for Scaling Up Inductive Algorithms." *Data Mining and Knowledge Discovery* 3 (2): 131-169, 1999.

54. Provost, F. and A. Danyluk, "Problem Definition, Data Cleaning and Evaluation: A Classifier Learning Case Study." *Informatica* 23: 123-136, 1999.

55. Fawcett, T. and F. Provost, "Adaptive Fraud Detection." *Data Mining and Knowledge Discovery* 1 (3):291-316, 1997.

56. Provost, F. and J. Aronis, "Scaling Up Machine Learning with Massive Parallelism." *Machine Learning* 23 (1):33-46, 1996.

57. Provost, F. and B. Buchanan, "Inductive Policy: The Pragmatics of Bias Selection." *Machine Learning* 20 (1/2):35-61, 1995

58. Krenzelok, E. and F. Provost, "The Ten Most Common Plant Exposures Reported to Poison Information Centers in the United States." *Journal of Natural Toxins* 4 (2): 195-202, 1995.

59. Provost, F. and R. Melhem, "A Distributed Algorithm for Embedding Trees in Hypercubes with Modifications for Run-Time Fault Tolerance." *Journal of Parallel and Distributed Computing* 14 (1): 85-89, 1992.

60. Clearwater, S., W. Cleland, F. Provost, E. Stern and Z. Zhang, "A Real-Time Expert System for Trigger Logic Monitoring." *Nuclear Instruments and Methods in Physics Research* A293: 491-495, 1990.

OTHER JOURNAL ARTICLES

61. Provost, F., Invited comment on Bolton and Hand's "Statistical Fraud Detection." *Statistical Science* 17 (2003).

62. "Applications of Data Mining to Electronic Commerce." Kohavi, R. and F. Provost. Guest editorial in *Data Mining and Knowledge Discovery* 5 (1/2) 2001.

63. "On Applied Research in Machine Learning." Provost, F. and R. Kohavi. Guest editorial in *Machine Learning* 30 (2/3) 1998.

SELECTED CONFERENCE PAPERS (MAINLY HIGHLY SELECTIVE CONFERENCES)

* A star indicates that the content of a paper is included completely or to some extent in a previously mentioned published paper, as detailed in braces following each case.

64. *"A Comparison of Methods for Treatment Assignment with an Application to Playlist Generation." Fernández-Loría, C., Provost, F., Anderton, J., Carterette, B., and Chandar, P. Workshop on Information Systems and Economics (**WISE**) 2020. **Best Student Paper Award Runner-up**.

65. "Combining Observational and Experimental Data to Improve Large-Scale Decision-Making." C. Fernandez & F. Provost. International Conference on Information Systems (**ICIS**), 2020.

66. *"Counterfactual Explanations for Data-Driven Decisions." C. Fernandez, F. Provost, X. Han. International Conference on Information Systems (**ICIS**), 2019.

67. "Iteratively Refining SVMs using Priors." E. Junque de Fortuny, T. Evgeniou, F. Provost, and D. Martens. *IEEE International Conference on Big Data (IEEE BigData 2015)*.

68. "Measuring causal impact of online actions via natural experiments: application to display advertising." D. Hill, R. Moakler, A. Hubbard, V. Tsemekhman, F. Provost, K, Tsemekhman. In the Proceedings of the Twenty-first ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD 2015**).

69. "Corporate Residence Fraud Detection." D. Martens, et al. In *Proceedings of the Twentieth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2014**).

70. "Scalable Hands-Free Transfer Learning for Online Advertising." Dalessandro, Brian, Daizhuo Chen, Troy Raeder, Claudia Perlich, Melinda Han Williams, and Foster Provost.. In *Proceedings of the Twentieth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2014**), pp. 1573-1582.

71. "Using Co-visitation Networks For Classifying Non-Intentional Traffic." Stitelman, O., C. Perlich, B. Dalessandro, R. Hook, T. Raeder, and F. Provost. In *Proceedings of the Nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2013**).

72. "Scalable Supervised Dimensionality Reduction Using Clustering." Raeder, T., C. Perlich, B. Dalessandro, O. Stitelman, and F. Provost. In *Proceedings of the Nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2013**).

73. "Bid Optimizing and Inventory Scoring in Targeted Online Advertising." Perlich, C., B. Dalessandro, R. Hook, O. Stitelman, T. Raeder, and F. Provost. In *Proceedings of the Eighteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2012**). **Best Paper Award**, Industry & Government Track.

74. "Design Principles of Massive, Robust Prediction Systems." Raeder, T., O. Stitelman, B. Dalessandro, C. Perlich, and F. Provost. In *Proceedings of the Eighteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2012**).

75. "Online Active Inference and Learning." Attenberg, J. and F. Provost. In *Proceedings of the Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2011**).

76. "Why label when you can search? Alternatives to active learning for applying human resources to build classification models under extreme class imbalance." Attenberg, J. and F. Provost. In *Proceedings of the Sixteenth ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining* (**KDD 2010**).

77. Attenberg, J. P. Melville and F. Provost. "A Unified Approach to Active Dual Supervision." In *Proceedings of the European Conference on Machine Learning and Principles of Knowledge Discovery in Databases* (**ECML PKDD 2010**).

78. Provost, F., B. Dalessandro, R. Hook, X. Zhang & A. Murray. Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting. In *Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD-2009)**.

79. *Sheng, S., F. Provost, and P. Ipeirotis. "Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers." In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD-2008**). **Best Paper Award Runner-up.**

80. Provost, F., P. Melville, and M. Saar-Tsechansky. "Data acquisition and cost-effective predictive modeling: targeting offers for electronic commerce." In *Proceedings of the Ninth International Conference on Electronic Commerce*, August 2007.

81. Melville, P., M. Saar-Tsechansky, F. Provost, and R. Mooney, "An Expected Utility Approach to Active Feature-value Acquisition." *Proceedings of the Fifth IEEE International Conference on Data Mining* (**ICDM-2005**), pp. 483-486. [short paper]

82. Macskassy, S., F. Provost, S. Rosset. "ROC Confidence Bands: An Empirical Evaluation." In *Proceedings of the 22nd International Conference on Machine Learning* (**ICML-2005**). [acceptance rate: 27%] [Also appears in the ICML-2005 Workshop on ROC Analysis in Machine Learning (ROCML-2005).]

83. Macskassy, S. and F. Provost. "Suspicion scoring based on guilt-by-association, collective inference, and focused data access." In Proceedings of the 2005 International Conference on Intelligence Analysis **(IA '05)**. [acceptance rate: 25%]

84. Melville, P., M. Saar-Tsechansky, F. Provost, and R. Mooney, "Active Feature-Value Acquisition for Classifier Induction." *Proceedings of the Fourth IEEE International Conference on Data Mining* (**ICDM-2004**), pp. 483-486. [short paper]

85. *Perlich, C., and F. Provost. "Aggregation-based Feature Invention and Relational Concept Classes." In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD-2003**). [acceptance rate: 13%] {*The second half of this paper is extended in paper 43.*}

86. Macskassy, S., H. Hirsh, F. Provost, R. Sankaranarayanan, V. Dhar. "Intelligent Information Triage." In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR-2001**)*. [acceptance rate: 23%]

87. *Saar-Tsechansky, M. and F. Provost. "Active Learning for Class Probability Estimation and Ranking." In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (**IJCAI-01**)*. [acceptance rate: 24%] {*Much of the work reported in this paper is included in paper 45.*}

88. Provost, F., D. Jensen and T. Oates, "Efficient Progressive Sampling." *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD-99**)*. [acceptance rate: 9%]

89. Fawcett, T. and F. Provost, "Activity Monitoring: Noticing Interesting Changes in Behavior." *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD-99**)*, pp. 53-62. DOI:https://doi.org/10.1145/312129.312195 [acceptance rate: 9%]

90. *Provost, F. and T. Fawcett, "Robust Classification Systems for Imprecise Environments." In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*. [acceptance rate: 30%] {*The work reported in this paper is included in paper 51.*}

91. Provost, F., T. Fawcett, and R. Kohavi "The Case Against Accuracy Estimation for Comparing Classifiers." *Proceedings of the Fifteenth Intl. Conference on Machine Learning (ICML-98).* [acceptance rate: 31%]

92. *Provost, F. and T. Fawcett, "Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions." In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*. **Best Paper Award Winner.** [acceptance rate: 10%] {*The work reported in this paper is included in paper 51.*}

93. *Provost, F. and V. Kolluri, "Scaling Up Inductive Algorithms: An Overview." In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*. [short paper; total acceptance rate (long+short): 41%] {*The work reported in this paper is included in paper 53.*}

94. Aronis, J. and F. Provost, "Increasing the Efficiency of Inductive Learning with Breadth-first Marker Propagation." In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*. [short paper; total acceptance rate (long+short): 41%]

95. Provost, F. and D. Hennessy, "Scaling Up: Distributed Machine Learning with Cooperation." In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*. [acceptance rate: 31%]

96. *Fawcett, T. and F. Provost, "Combining Data Mining and Machine Learning for Effective User Profiling." In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. [acceptance rate: 20%] {*The work reported in this paper is included in paper 55.*}

97. Aronis, J., F. Provost, and B. Buchanan, "Exploiting Background Knowledge in Automated Discovery." In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. [short paper; total acceptance rate (long+short): 33%]

98. Provost, F., "Iterative Weakening: Optimal and Near-Optimal Policies for the Selection of Search Bias." In *Proceedings of the Eleventh National Conf. on Artificial Intelligence (AAAI-93)*. [acceptance rate: 25%]

99. Danyluk, A. and F. Provost, "Small Disjuncts in Action: Learning to Diagnose Errors in the Telephone Network Local Loop." In *Proceedings of the Tenth International Conference on Machine Learning (ICML-93)*. [acceptance rate: 27%]

100. *Provost, F. and B. Buchanan, "Inductive Policy." In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*. [acceptance rate: 21%] {*Some of the work reported in this paper is included in paper 57.*}

ADDITIONAL CONFERENCE PAPERS

101. "Revealing Life Events from Inferred Customer Similarity: A Predictive Modeling Approach." T. Nian, Junqué de Fortuny, F. Provost. Proceedings of WITS 2014. [acceptance rate: 38%]

102. "Hyperlocal: Inferring Location of IP Addresses in Real-time Bid Requests for Mobile Ads." L. Long, T. Eliassi-Rad, F. Provost, and L. Moores. In Proceedings of the 6ᵗʰ ACM SIGSPATIAL International Workshop on Location-based Social Networks (LBSN 2013).

103. *Macskassy, S. and F. Provost. "Suspicion scoring of networked entities based on guilt-by-association, collective inference, and focused data access." In Annual Conference of the North American Association for Computational Social and Organizational Science (NAACSOS), 2005. [This is a followup paper to paper 83, with new results but considerable overlap.]

104. Macskassy, S. and F. Provost. "NetKit-SRL: A Network Learning Toolkit and its use for classification of networked data." In Annual Conference of the North American Association for Computational Social and Organizational Science (NAACSOS), 2005.

105. Kolluri, V., F. Provost, B. Buchanan, and D. Metzler. "Knowledge Discovery Using Concept-Class Taxonomies." In *Lecture Notes in Computer Science* 3339. *AI 2004: Advances in Artificial Intelligence: 17th Australian Joint Conference on Artificial Intelligence*. Springer-Verlag Heidelberg.

106. Provost, F. and D. Hennessy, "Distributed Machine Learning: Scaling up with Coarse-grained Parallelism." In *Proc of the Second International Conference on Intelligent Systems for Molecular Biology (ISMB-94)*.

107. Provost, F., "ClimBS: Searching the Bias Space." In *Proceedings of the Fourth International IEEE Conference on Tools with Artificial Intelligence (TAI-92)*. [acceptance rate: 36%]

108. Provost, F. and R. Melhem, "Embedding Rings in Hypercubes for Run-Time Fault Tolerance." In *Proceedings of the Fourth ISMM/IASTED Intl. Conference on Parallel and Distributed Computing and Systems*, 1991.

109. Clearwater, S. and F. Provost, "RL4: A Tool for Knowledge-Based Induction." In *Proceedings of the Second International IEEE Conference on Tools for Artificial Intelligence (TAI-90)*. [acceptance rate: 40%]

OTHER WORKSHOP/SYMPOSIUM PAPERS

110. "Detecting Employee Misconduct and Malfeasance" George Valkanas, Panagiotis Ipeirotis, Foster Provost, Josh Attenberg, Jennifer Chin, Chathra Hendahewa, Abe Stanway, Bernando Suryanto and Bharath Vivekananda Swamy. *MIS2: Misinformation and Misbehavior Mining on the Web*, 2018, Marina Del Rey, CA, USA.

111. "Enhancing Transparency and Control when Drawing Data-Driven Inferences about Individuals." Daizhuo Chen, Samuel P. Fraiberger, Robert Moakler, Foster Provost. https://arxiv.org/abs/1606.08063

112. "Explaining Classification Models Built on High-Dimensional Sparse Data." Julie Moeyersoms, Brian d'Alessandro, Foster Provost, David Martens. ICML-2016 Workshop on Human Interpretability in Machine Learning (WHI 2016). https://arxiv.org/abs/1607.06280

113. "Pleasing the advertising oracle. Probabilistic prediction from sampled, aggregated ground truth." M. Williams, C. Perlich, B. Dalessandro, F. Provost. In Proceedings of the Eighth International Workshop on Data Mining for Online Advertising (ADKDD 2014).

114. Moakler, R., et al. "Causal impact of online advertisements using viewability as a method of treatment." Winter Conference on Business Intelligence, Feb. 2014.

115. "Estimating Audience Interest Distributions based on Audience Web Behavior." X. Zhang, F. Provost, K. Tsemekhman. Winter Business Intelligence Conference, 2012.

116. "Causally Motivated Attribution for Online Advertising." B. Dalessandro, O. Stitelman, C. Perlich, and F. Provost. *ADKDD* 2012.

117. "Geo-social Network Advertising." F. Provost, D. Martens and A. Murray. Winter Business Intelligence Conference, 2012.

118. "Estimating the Effect of Online Display Advertising on Browser Conversion." O. Stitelman, B. Dalessandro, C. Perlich, and F. Provost. *ADKDD* 2011.

119. "Beat the Machine: Challenging crowdsourcing workers to find the unknown unknowns." Attenberg, J., P. Ipeirotis & F. Provost. *Proceedings of the Third Human Computation Workshop (HCOMP 2011)*.

120. "Managing Crowdsourcing Workers." with Jing Wang and Panos Ipeirotis. Winter Business Intelligence Conference, 2011.

121. "Quality management on Amazon Mechanical Turk," P. Ipeirotis, F. Provost, and J. Wang. *Proceedings of the Second Human Computation Workshop (KDD-HCOMP 2010)*.

122. Attenberg, J. and F. Provost. "Active inference and learning for classifying streams." In *ICML-2010 Workshop on Budgeted Learning*.

123. Attenberg, J., P. Melville and F. Provost. "Guided feature labeling for budget-sensitive learning under extreme class imbalance." *In ICML-2010 Workshop on Budgeted Learning*.

124. Zheng, R., F. Provost, and A. Ghose. "Social Network Collaborative Filtering: Preliminary Results." The Sixth Workshop on e-Business (WeB 2007).

125. Hill, S., F. Provost and C. Volinsky (2007). "Learning and Inference in Massive Social Networks." The 5th International Workshop on Mining and Learning with Graphs, August 2007. Extended Abstract.

126. *Macskassy, S, and F. Provost (2006). "A brief survey of machine learning methods for classification in networked data and an application to suspicion scoring." E.M. Airoldi et al. (Eds.): ICML 2006 Ws, *LNCS* **4503**, pp. 172-175. Springer-Verlag. (Originally appeared at the Workshop on Statistical Network Learning at 23rd International Conference on Machine Learning (ICML 2006), Pittsburgh, 29 June 2006.) {There is considerable overlap between this paper and paper 40.}

127. *Melville, P., M. Saar-Tsechansky, F. Provost, R. Mooney. "Economical Active Feature-value Acquisition through Expected Utility Estimation." In the KDD-05 Workshop on Utility-based Data Mining. {There is considerable overlap between this paper and paper 64.}

128. Giraud-Carrier, C. and F. Provost. "Toward a Justification of Meta-learning: Is the No Free Lunch Theorem a Show-stopper?" In the ICML-2005 Workshop on Meta-Learning.

129. Macskassy, S., F. Provost, S. Rosset. "Pointwise ROC Confidence Bounds: An Empirical Evaluation." In the ICML-2005 Workshop on ROC Analysis in Machine Learning (ROCML-2005).

130. Hill, S., F. Provost and C. Volinsky. "The Gift of Gab: Evidence TelE-Commerce Firms can Profit from Viral Marketing." (Two-page extended abstract) First Interdisciplinary Symposium between Information Systems, Statistics and Related Fields. Decision and Information Technologies Department, Robert H. Smith School of Business, Univ. of Maryland. May 2005.

131. Macskassy, S. and F. Provost. "Confidence Bands for ROC Curves: Methods and an Empirical Study." First Workshop on ROC Analysis in AI. August 2004.

132. Macskassy, S. and F. Provost. "A Simple Relational Classifier." In the Second Workshop on Multi-Relational Data Mining (MRDM-2003) at KDD-2003.

133. Bernstein, A., S. Clearwater, and F. Provost. "The Relational Vector-space Model and Industry Classification." To appear in *Proceedings of the IJCAI-2003 Workshop on Learning Statistical Models from Relational Data*.

134. *Provost, F., C. Perlich, and S. Macskassy. "Relational Learning Problems and Simple Models." To appear in *Proceedings of the IJCAI-2003 Workshop on Learning Statistical Models from Relational Data*. {*Some of the work reported in this position paper is included in paper 0.*}

135. *Perlich, C. and F. Provost. "Aggregation and Concept Complexity in Relational Learning." To appear in *Proceedings of the IJCAI-2003 Workshop on Learning Statistical Models from Relational Data*. {*Some of the work reported in this position paper is included in paper 64.*}

136. Bernstein, A., S. Clearwater, S. Hill, C. Perlich, and F. Provost. "Discovering Knowledge from Relational Data Extracted from Business News." In *Proceedings of the KDD-2002 Workshop on Multi-Relational Data Mining*, 2002.

137. *Bernstein, A. and F. Provost. "An Intelligent Assistant for the Knowledge Discovery Process." In Proceedings of IJCAI-01 Workshop on Wrappers for Performance Enhancement in Knowledge Discovery in Databases, 2001. {*The work reported in this paper is included in working paper 7.*}

138. *Macskassy, S., H. Hirsh, F. Provost, R. Sankaranarayanan, V. Dhar. Information Triage using Prospective Criteria. In *Proceedings of the Eighth International Conference on User Modeling (UM-2001), Workshop on Machine Learning, Information Retrieval and User Modeling*, 2001. {*The work reported in this paper also appears in paper 86.*}

139. Aronis, J., V. Kolluri, F. Provost, and B. Buchanan, "The WoRLD: Knowledge Discovery from Multiple Distributed Databases." *In Proc. of the Florida Artificial Intelligence Research Symposium (FLAIRS-97)*.

140. Provost, F. and A. Danyluk, "Learning from Bad Data." In *Proceedings of the ML-95 Workshop on Applying Machine Learning in Practice*, 1995.

141. Krenzelok, E., F. Provost, T. Jacobsen, J. Aronis, B. Buchanan, "Assessing Patient Referral Patterns to a Health Care Facility in Plant Exposure Patients Using Computer Artificial Intelligence." European Assn. of Poison Centres and Clinical Toxicologists Scientific Meeting. May 18-20, 1995, Krakow, Poland.

142. Krenzelok, E., F. Provost, T. Jacobsen, J. Aronis, B. Buchanan, "Poinsettia (Euphorbia pulcherrima) Exposures Have Good Outcomes...Just As We Thought." European Association of Poison Centres and Clinical Toxicologists Scientific Meeting, 1995.

143. Aronis, J. and F. Provost, "Efficiently Constructing Relational Features from Background Knowledge for Inductive Machine Learning" In *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases*, (**KDD-94**).

144. Provost, F., "Goal-Directed Inductive Learning: Trading Off Accuracy for Reduced Error Cost." In *Proceedings of the AAAI Spring Symposium on Goal-Directed Learning*, 1994.

145. Danyluk, A. and F. Provost, "Adaptive Expert Systems: Applying Machine Learning to NYNEX MAX." In *Proceedings of the AAAI-93 Workshop: AI in Service and Support--Bridging the Gap between Research and Applications*, 1993.

146. Provost, F., "A Baseline Taxonomy of Bias Adjustment Policies." In *Proceedings of the ML-92 Workshop on Biases in Learning*, 1992.

BOOK CHAPTERS & OTHER PUBLICATIONS

147. "Selective Data Acquisition for Machine Learning." Attenberg, J., P. Melville, F. Provost and M. Saar-Tsechansky. In B. Krishnapuram, S. Yu, and B. Rao (Eds.), Cost-Sensitive Machine Learning. Chapman & Hall/CRC, 2011.

148. Perlich, C. and F. Provost. Aggregation for Predictive Modeling with Relational Data." *Encyclopedia of Data Warehousing and Mining*, 2005.

149. *Provost, F. and V. Kolluri, "Scalability." In W. Kloesgen and J. Zytkow (eds*.), Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, 2002, pp. 418-433. {*Some of the work reported in this paper is included in paper 53*.}

150. Danyluk, A. and F. Provost, "Telecommunications Network Diagnosis." In W. Kloesgen and J. Zytkow (eds.), *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, 2002, pp. 897-902.

151. *Fawcett, T. and F. Provost, "Fraud Detection." W. Kloesgen and J. Zytkow (eds.), *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, 2002, pp. 726-731. {*Some of the work reported in this paper is included in paper 55*.}

152. Provost, F. "Learning with Imbalanced Data Sets 101." Invited paper for the AAAI-2000 Workshop on Learning with Imbalanced Data Sets, 2000.

153. *Provost, F., D. Jensen and T. Oates, "Progressive Sampling." In H. Liu and H. Motoda (eds*.), Instance Selection and Construction, A Data Mining Perspective*, Kluwer Academic Publishers, 2000. {*The work reported in this paper is included in paper 88*.}

154. *Provost, F., "Distributed Data Mining: Scaling up and beyond." In H. Kargupta and P. Chan (eds.), *Advances in Distributed Data Mining*, San Francisco, CA: Morgan Kaufmann, 2000. {*Some of the work reported in this paper is included in paper 53*.}

155. Danyluk, A., T. Fawcett, and F. Provost, "AI Approaches to Time-series Problems." Workshop report in *AI Magazine*, 1999.

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

157. Provost, F. and B. Buchanan, "Inductive Strengthening: The effects of a simple heuristic for restricting hypothesis space search." In K.P. Jantke (ed.), *Analogical and Inductive Inference (Lecture Notes in Artificial Intelligence* 642*)*. Springer-Verlag, 1992.

158. *Clearwater, S., W. Cleland, F. Provost, E. Stern and Z. Zhang, "A Real-Time Expert System for Experimental High Energy/Nuclear Physics." In D. Perrett-Gallix and W. Wojcik (eds.), *New Computing Techniques in Physics Research*. Paris: Centre National de la Recherche Scientific, 1990. {*The work reported in this paper is included in paper 60*.}

159. *Provost F. and R. Melhem, "Distributed Fault Tolerant Embedding of Trees and Rings in Hypercubes." In I. Koren (ed.) *Defect and Fault Tolerance in VLSI systems Volume 1*. New York, NY: Plenum Press, 1989. {*Some of the work reported in this paper is included in paper 59 and some is reported in paper 108*.}

OTHER WORKING PAPERS

160. "Social Network Collaborative Filtering." Zheng, R., D. Wilkinson, and F. Provost. Working paper CeDER-8-08. Center for Digital Economy Research, Stern School of Business, New York University. September 2008.

161. *Provost, F. and P. Domingos. "Well-trained PETs: Improving Probability Estimation Trees" CeDER Working Paper #IS-00-04. {*Early version of paper 50*.}

162. Provost, F., J. Aronis, and B. Buchanan. "Rule-space Search for Knowledge-based Discovery." CIIO Working Paper #IS 99-012.

SELECTED TUTORIALS

"Data Science for Real Estate." Tutorial at the ACM SIGKDD International Conference on Data Mining and Knowledge Discovery, August 2020.

"Predictive Modeling with Social Networks." Tutorial (invited) at the Conference on Complex Networks: from theory to interdisciplinary applications, July 2016.

"Predictive Modeling with Social Networks." Tutorial (invited) at the Third International AAAI Conference on Weblogs and Social Media (with Jennifer Neville), May 2009.

"Predictive Modeling with Social Networks." Tutorial at the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (with Jennifer Neville), August 2008.

"Social Network Mining: A Tutorial on Inference and Learning with Social Network Data." Tutorial (invited) at the Twenty-Third AAAI Conf. on Artificial Intelligence (with Jennifer Neville), July 2008.

"Inference and Learning with Networked Data." Invited tutorial lecture for the NATO Advanced Study Institute on Mining Massive Data Sets for Security, September 2007.

"Modeling Complex Networks for Electronic Commerce." Tutorial at the ACM Conference on Electronic Commerce (with Arun Sundararajan), June 2007.

"Evaluating Machine Learning, Knowledge Discovery, and Data Mining" (with David Jensen). Tutorial presented at **KDD-98**, at **AAAI-99** and at **IJCAI-99**.

SELECTED INVITED TALKS (Etc.)

2022 Academic Talks:
    RPI, Feb 2022
    Pitt, Feb 2022

"Causal Targeting - It's not Causal Effect Estimation (and Why it Matters)." Keynote lecture for IEEE International Conference on Data Science and Advanced Analytics, October 2021.

"A Decade of Human Computation." Keynote Fireside Chat (panel) for the ACM Conference on Human Computation (in celebration of the 10th anniversary of the conference), October 2019.

"The Predictive Power of Data about our Fine-Grained Behavior." Keynote Lecture for Data Science for Business in the 21st Century. University of Antwerp, October 2019.

"Can Bad Guys Hide from the Network?" Invited talk for Complex Networks: from socio-economic systems to biology and brain, July 2019.

"AI, Machine Learning, Data Science, Prediction & Risk." Invited talk for the InsurTech Alliance NYC meeting, May 2019.

Invited Panel Presentation, Innovative Practices on Fraud Detection and Prevention, NYU Law School Program on Corporate Compliance and Enforcement, April 2019.

"Beyond the Code: Success in analytics requires more than just technical skills." Keynote lecture for BAC2, NYU|Stern Business Analytics Club's Business Analytics Conference, Spring 2019.

"The Predictive Power of Data about our Fine-Grained Behavior." Keynote Lecture for OneAnalyst Conference, Jan 2019, Microsoft, Bellevue, WA.

"Explaining the Decisions Made by Machine Learning- based Models." Keynote Lecture for 2018 Conference on Digital Marketing and Machine Learning, Dec 2018, Carnegie Mellon University.

"Technology & Data-Driven Compliance." Panel presentation at NYU Law School Compliance Conference, Oct 2018.

"The Predictive Power of Data on Consumers' Fine-Grained Behavior." Invited talk at IEEE DSAA 2018 (request from conference sponsor).

2018 Academic Talks:
- HBS (Feb)
- Boston University (March)
- Minnesota (March)
- IIT (April)
- NYU's Music Business Program (September)
- MIT (September)

"The Predictive Power of Massive Data about our Fine-Grained Behavior." Invited Lecture at MIT Sloan School, September 2017.

"So You've Built a Machine Learning Model…Now What?" Invited talk for AT&T Research, Nov. 2017.

Invited Panel Presentation at the INFORMS Data Science meeting, Oct. 2017. Panel: Publishing Data Science Papers in IS Journals: Challenges and Editorial Advice

Invited Panel Presentation at the MIS Leadership Conference at U. Maryland, Nov. 2017. Panel: Industry-Academia Collaboration and Research Grants

Invited Panel Presentation at the annual meeting of the Institute for Scientific Interchange Foundation, Turin Italy, October 2017. (On Data Science, Bias, Discrimination, and Ethics.)

"Using Data Analytics to Enhance Compliance." Invited panel presentation for Compliance: New Risks, Challenges, and Approaches, Conference held by NYU Program on Corporate Compliance and Enforcement, October 2016.

"The Predictive Power of Massive Data about our Fine-grained Behavior." Keynote talk for the 15th International Symposium on Intelligent Data Analysis (IDA 2016), October 2016.

"Data-driven Models: Can we explain why they do what they do?" Invited talk for *Science Crossroads*, annual meeting of the Institute for Scientific Interchange Foundation, Turin Italy, October 2016.

"Competing with Analytics in the Connected Economy." Keynote talk for NYU/Stern's Business Analytics Symposium, September 2016.

"Data Science for Connected Commerce." Invited Lecture at MIT Sloan School, September 2016.

"Academia & Entrepreneurship." Invited talk for Univ. of Alberta's Machine Intelligence Institute, Sept. 2016.

"The Predictive Power of Massive Data about our Fine-grained Behavior." Keynote talk for the 9th ACM International Conference on Web Search and Data Mining (WSDM), February 2016.

"Privacy, Confidentiality, Transparency & Control: From a fundamental tension in data science to the 'explanation trick.'" Invited panel talk & discussion, FATML workshop at NIPS 2014.

"Teaching Data Science." Invited panel on Data Science Education, Palo Alto, December 2014.

"Understanding Decisions Driven by Big Data: From Analytics Management to Privacy-friendly Cloaking Devices." Keynote lecture at StrataEU, November 2014.

"Predictive Analytics with Fine-grained Behavior Data." Invited talk, Data-driven Business Day, StrataEU, November 2014.

"Data Science for Business." Invited Lecture at MIT Sloan School, October 2014.

"Whiskey Analytics." Lecture and whiskey tasting, inaugural event for Stern MBA Business Analytics Club, September 2014.

"A Conversation with Foster Provost." NYU President's Council & NYU Giving Society Event, May 2014.

"Predictive Models from Massive Fine-Grained Data Work Great, But Do You Understand What They're Doing?" Invited talk for MIT's Workshop on Big Data Analytics, March 2014

"Display Ad Targeting based on Fine-Grained Consumer Behavior." Invited talk at the 2013 Marketing in Israel Conference, December 2013.

"Data Science for Business." Invited lecture (based on the book) at the Institute for Big Data Analytics at Dalhousie University, Nov 2013.

"Big Data: Is Bigger Really Better? Predictive Modeling with Fine-Grained Behavior Data." Keynote talk at Strata NY, Oct 2013

"Mining (Massive) Consumer Behavior Data for Predictive Marketing." Invited talk at CMU, Nov. 2013.
"Mining (Massive) Consumer Behavior Data for Predictive Marketing." Invited talk at Rutgers U., May 2013.
"Mining (Massive) Consumer Behavior Data for Predictive Marketing." Invited talk at the Microsoft Research Machine Learning Summit, Paris, April 2013.

"Mining (Massive) Consumer Behavior Data for Marketing." Keynote talk at the IEEE International Conference on Data Mining (ICDM), Industry Track, Dec 2012.
"Mining (Massive) Consumer Behavior Data for Marketing." Invited talk at IBM Research, Oct 2012.

"Social Targeting for Privacy-Friendly On-line Advertising." Invited talk at University of Antwerp, Nov. 2011.

"Social Targeting for Privacy-Friendly On-line Advertising." Invited talk for the 2011 International Conference on Weblogs and Social Media, July 2011.

"Machine Learning for Display Advertising." Keynote lecture for the 2010 Workshop on Machine Learning for Online Advertising, December 2010.

"(Privacy-friendly!) Social Network Targeting for On-line Advertising" Invited talk for KDD-2010.

"Get Another Label? Improving Data Quality and Machine Learning using Multiple, Noisy Labelers." Invited talk for NYU/Poly, November 2010.

"Get Another Label? Improving Data Quality and Machine Learning using Multiple, Noisy Labelers." Invited talk for MIT Operations Research Center, November 2010.

"Get Another Label? Improving Data Quality and Machine Learning using Multiple, Noisy Labelers." Invited talk for Columbia University Center for Computational Learning Systems, April 2010.

"Social Targeting for Privacy Friendly Online Advertising." Invited talk for NYU/Stern/L2 Social Commerce Clinic, October 2010.

"Social-Media-Based Targeting for Privacy-Friendly On-Line Advertising." Edison Lecture, Notre Dame University, October 2010.

"Active Inference and Learning for Data Streams." Invited Talk for INFORMS, November 2010.

"Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting." Invited talk for Wharton WIMI Conference on the Emergence and Impact of User-Generated Content, December 2009.

"Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting." Invited talk for Tufts University, October 2009.

"Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting." Invited talk at the INFORMS annual meeting, October 2009.

"Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting." Invited talk at the Conference on Marketing meets Data Mining, August, 2009.

"Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting." Invited talk at the Third International Workshop on Data Mining and Audience Intelligence for Advertising, June 2009.

"Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting." Invited talk for Carnegie Mellon University, April 2009.

"Inference and Learning with (Social) Network Data." Invited talk for Williams College, April 2008.

"Inference and Learning with Networked Data." Invited talk for KXEN Corp. Paris, January 2008.

"Inference and Learning with Networked Data." Invited talk reprise, Dept. of Applied Economic Sciences, Katholieke Universiteit Leuven, January 2008.

"Inference and Learning with Networked Data." Invited talk, Univ. Conn. Business School, November 2007

"Inference and Learning with Networked Data." Invited talk, Dept. of Applied Economic Sciences, Katholieke Universiteit Leuven, September 2007.

"Inference and Learning with Networked Data." Invited talk (one of two) for the Sixth Annual Computer Science Day, Fordham University, April 2007.

"Inference and Learning with Networked Data." Invited talk, INSEAD, March 2007.

"Classification in Networked Data." Invited talk, University of Zurich, November 2006.

"Mining Social Networks for Fun and Profit." Invited talk, University of Pittsburgh, October 2006.

"Data Mining for Business Intelligence." Invited talk for New York CTO Club, October 2005.

"Economic Machine Learning and Utility-based Data Mining" Invited talk (one of three) KDD-2005 Workshop on Utility-based Data Mining.

"Toward a Justification of Meta-learning: Is the No Free Lunch Theorem a Show-stopper?" Invited talk (one of one) ICML-2005 Workshop on Meta-Learning.

"Classification and Learning with Networked Data" University of Notre Dame. April 2005.

"Classification and Learning with Networked Data" Brigham Young University. March 2005.

"Classification and Learning with Networked Data" Google Research. February 2005.

"Defining the Research Agenda for Business Intelligence." Invited panel presentation for the 2005 Utah Winter Information Systems Conference. David Eccles School of Business, University of Utah. March 2005.

"Classification and Learning with Networked Data: Some observations and results." Invited talk for the Red McComb's School of Business and the Department of Computer Science, University of Texas at Austin, November 2004.

"Active Information Acquisition for Machine Learning." Invited talk for Amazon.com's data mining group, August 2004.

"Obtaining External Funding." Invited talk for the Ph.D. Project in Information Systems Conference, August 2004.

"The role of applications in the science of machine learning." Invited talk (one of three) for the Twentieth International Conference on Machine Learning, August 2003.

"Marginal class distribution and budget-sensitive learning." Invited talk (one of three) for the ICML-2003 Workshop on Learning from Unbalanced Data Sets, August 2003.

"Research in Data Mining and Machine Learning." Invited talk for the Ph.D. Project in Information Systems Conference, August 2003.

"The Machine Learning Journal." Invited panel presentation for the Ph.D. Project in Information Systems Conference (journal editors panel), August 2003.

"Data mining and cross-selling." Invited talk for Pitney Bowes' Advanced Concepts and Technology division, July 2003.

"Data mining and business intelligence." Invited talk for Pitney Bowes' Advanced Concepts and Technology division, January 2003.

"Adaptive Fraud Detection." Invited talk for the Rutgers University Electronic Commerce Association, December 2002.

"Tree induction versus logistic regression." Keynote talk (one of four) at SAS's annual Data Mining Conference, October 2002.

"Tree induction versus logistic regression." Invited talk at the annual meeting of the Classification Society of North America, June 2002.

"Adaptive Fraud Detection." Invited talk at Stevens Institute of Technology, April 2002.

"Intelligent Assistance for the Knowledge Discovery Process." Invited talk at INFORMS, Nov. 2001.

"Reviving the Learning Curve: A critical study of machine learning performance." Invited talk at the University of Texas at Austin, Data Mining Lecture Series, Oct. 2001.

"Wrappers: Intelligent Support for the Knowledge Discovery Process." Invited talk (one of two) for the IJCAI-01 Workshop on Wrappers for Performance Enhancement in KDD, August 4, 2001.

"Machine Learning for Information Monitoring and Triage." Invited talk (one of three) for the 14th Canadian Conference on Artificial Intelligence, June 2001.

"Monitoring and Triage of Business News." Invited talk for the DIMACS Working Group on Data Analysis and Digital Libraries, DIMACS Center, Rutgers University, May 17, 2001.

"Robust Classification for Imprecise Environments." Invited presentation at INFORMS, Nov. 2000.

"Information Monitoring and Triage for Mobile Alerting and Timely Decision Making." Mathematical Sciences Department, IBM T.J. Watson Research Center, October, 2000.

"Tutorial Overview: The Cost-Sensitive Learning Problem(s) and a Review of the Main Existing Approaches to Solving Them." Invited (keynote) talk for the ICML Workshop on Cost-sensitive Learning, 2000.

"Learning with Imbalanced Data Sets 101." Invited talk for the AAAI Workshop on Learning with Imbalanced Data Sets, 2000.

"Well-behaved PETs: Improving Probability Estimation Trees." Invited presentation at University of Pittsburgh and at Washington State University, Fall 1999.

"Rolling the Dice with Data Mining Success." KDD-98 Workshop on the Keys to the Commercial Success of Data Mining, August 1998.

"Distributed Data Mining: Scaling up and beyond." KDD-98 Workshop on Distributed Data Mining, Aug 1998.

"Evaluating Data Mining Results." Annual Symposium of the ACM Central Ohio Chapter, 1998.

"How Well Does Data Mining Work? Evaluating the Knowledge Discovered." Third Annual Data Mining Summit, March 1998.

"Comparing Classifiers under Imprecise Business Criteria." Mathematical Sciences Department, IBM T.J. Watson Research Center, September 1997.

"Fraud Detection Systems: Combining Data Mining and Machine Learning." Second Annual Data Mining Summit, February 1997.

"Automated Design of Fraud Detection Systems." Brown University Industrial Partners Program, Nov. 1996.

"Pushing Machine Learning into the Real World." NYNEX Science and Technology, May 1994.

"What Do Learning Goals Look Like?" (Panel) AAAI Spring Symposium on Goal-Directed Learning, 1994.

"Machine Learning & 'Real-World' Complications." NYNEX Science and Technology, May 1993.

"Policies for Machine Learning." Mathematical Sciences Dept., IBM T.J.Watson Research Center, April 1993.

"Policies for Inductive Bias Selection." NYNEX Science and Technology, May 1992.