Foster Provost

PERSONAL DATA

Information Systems Group
Technology Operations and Statistics Department
Leonard N. Stern School of Business
New York University
44 West 4th Street, Suite 8-84
New York, NY 10012-1126, USA

Phone: +1 (212) 998-0806
Fax: +1 (212) 995-4228
Email: fprovost@stern.nyu.edu
Website: https://fosterprovost.com

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

PROFESSIONAL EMPLOYMENT

- 2018 - now: Director, Data Analytics & AI, Fubon Center, Stern School, NYU, New York, NY, USA
- 2015 - now: Professor of Data Science, New York University, New York City, NY, USA
- 2015 - 2016: Director (interim), Center for Data Science, NYU, New York City, NY, USA
- 2008 - now: Professor of Information Systems, Stern School of Business, New York University, New York City, NY, USA
- 2001 - 2008: Associate Professor of Information Systems, Stern School of Business, New York University, New York City, NY, USA
- 1999 - 2001: Assistant Professor of Information Systems, Stern School of Business, New York University, New York City, NY, USA
- 1994 - 1999: Research Scientist, NYNEX Science and Technology, Inc. (now Verizon), New York City, NY, USA

AWARDS AND GRANTS

- 2020: 2020 ACM SIGKDD Test of Time Award; Paper: "Get Another Label ..." with Sheng and Ipeirotis
- 2020: Best Student Paper Award Runner-up, WISE 2020 ; Paper: A Comparison of Methods for Treatment Assignment with an Application to Playlist Generation (Carlos Fernández-Loría was the student)
- 2018: Finalist, 2018 Algorithm for Change AI/ML Competition (Helping low income, underrepresented minority and first generation students get to and through college); Algorithm for Change AI/ML Competition
- 2017: European Research Paper of the Year; AIS & CIONET
- 2016: Best Paper Award; Information Systems Research
- 2015: 3 of the Top-5 Most Downloaded Papers (including #1); Big Data
- 2015: INFORMS President’s Pick, for paper published in Information Systems Research; INFORMS
- 2014: MSBA Best Teacher Award; NYU/Stern
- 2014: Nominated for 2014 NYU/Stern Professor of the Year by the MBA student body (one of 31 nominees); NYU/Stern
- 2014: Wharton Customer Analytics Initiative Awards; Wharton
- 2013: Nominated for 2013 NYU/Stern Professor of the Year by the MBA student body (one of 23 nominees); NYU/Stern
- 2012: Wharton Customer Analytics Initiative Awards; Wharton
- 2012: People’s Choice Award, Runner-up; Future of Advertising Conf. on Empirical Generalizations in Advertising
- 2012: Best Paper Award; ACM SIGKDD 2012, Industry Track

- 2011: Wharton Customer Analytics Initiative Awards; Wharton
- 2009: The INFORMS Design Science Award; INFORMS
- 2009: Marketing Science Institute Award; Marketing Science Institute
- 2009: Marketing Science Institute Award, “On-line brand advertising using social networks induced from user-generated content” , $6,500 Marketing Science Institute
- 2008: Best Paper Award Runner-up; ACM SIGKDD 2008
- 2008: IBM OCR Award, "Data and Risk", $200,000 IBM
- 2004: DOD/AFRL Grant; , $186,925 Evidence Assessment, Grouping, Linking, and Evaluation Program
- 2004: NYSIA Grant, "Algorithms and software for suspicion scoring", $35,000 NYSIA
- 2003: Third Place; KDDCUP-2003
- 2003: Second Place; KDDCUP-2003
- 2001: IBM Faculty Awards; IBM
- 2001: DARPA Grant; , $825,000 Evidence Extraction and Link Discover Program
- 2001: Elected as one of ten founding board members of the Intl. Machine Learning Society; Intl. Machine Learning Society
- 2000: IBM Faculty Awards, "Extending the Reach of Data Mining", $80,000 IBM
- 2000: IBM Faculty Awards; IBM
- 1997: Best Paper Award; Third International Conference on Knowledge Discovery and Data Mining (KDD-97)
- 1995: President's Award; NYNEX Science and Technology (now Verizon)
- 1994: Co-PI: NSF Grant NSF-IRI-9412549, "Using Parallelism to Scale Up Machine Learning to Large Data-Analysis Problems", $240,101 NSF
- 1986: Summa Cum Laude (4.0 GPA), and awarded the Certificates of General Excellence, Excellence in Mathematics, and Excellence in Physics; Duquesne University

TEACHING EXPERIENCE

- Fall 2018: Introduction to Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2018: Introduction to Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2018: Data Science for Biz Analytics (Technical) (MBA Class); Instructor; NYU/Stern
- Spring 2018: Data Mining for Business Analytics (Ugrad Class); Instructor; NYU/Stern
- Fall 2017: Introduction to Business Analytics (MSBA Class); Instructor; NYU/Stern
- Fall 2017: Data Mining for Biz Analytics (Managerial) (MBA Class); Instructor; NYU/Stern
- Spring 2017: Introduction to Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2017: Data Science for Biz Analytics (Technical) (MBA Class); Instructor; NYU/Stern
- Spring 2017: Data Science Research (Ph.D. Seminar); Instructor; NYU/Stern
- Fall 2016: Data Mining for Biz Analytics (Managerial) (MBA Class); Instructor; NYU/Stern
- Fall 2016: Data Analytics (TRIUM Exec Education Class); Instructor; NYU/Stern
- Fall 2016: Data Science for Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2016: Data Analytics (TRIUM Exec Education Class); Instructor; NYU/Stern
- Spring 2016: Data Science for Business Analytics (MSBA Class); Instructor; NYU/Stern
- Fall 2015: Data Analytics (TRIUM Exec Education Class); Instructor; NYU/Stern
- Fall 2015: Intro to Data Science/Data Mining for Biz Analytics (MSDS/MBA Class); Instructor; NYU/Stern
- Fall 2015: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern
- Spring 2015: Data Analytics (TRIUM Exec Education Class); Instructor; NYU/Stern
- Fall 2014: Data Science for Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2014: Data Science for Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2014: Data Science Research (Ph.D. Seminar); Instructor; NYU/Stern
- Fall 2013: Intro to Data Science/Data Mining for Biz Analytics (MSDS/MBA Class); Instructor; NYU/Stern
- Fall 2013: Data Science for Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2013: Data Science for Business Analytics (MSBA Class); Instructor; NYU/Stern
- Spring 2013: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern
- Spring 2013: Data Science Research (Ph.D. Seminar); Instructor; NYU/Stern
- Fall 2012: Practical Data Science (MBA Clas); Instructor; NYU/Stern
- Fall 2012: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern

- Spring 2012: Data Science Research (Ph.D. Seminar); Instructor; NYU/Stern
- Spring 2011: Data Mining for Business Analytics (Ugrad Class); Instructor; NYU/Stern
- Spring 2011: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern
- Spring 2011: Networks, Crowds & Markets (MBA/PhD Class); Instructor; NYU/Stern
- Spring 2010: Data Mining for Business Analytics (Ugrad Class); Instructor; NYU/Stern
- Spring 2010: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern
- Spring 2009: Data Mining for Business Analytics (Ugrad Class); Instructor; NYU/Stern
- Spring 2009: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern
- Spring 2008: Data Mining for Business Analytics (Ugrad Class); Instructor; NYU/Stern
- Spring 2008: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern
- Spring 2007: Data Mining for Business Analytics (Ugrad Class); Instructor; NYU/Stern
- Spring 2007: Data Mining for Business Analytics (MBA Class); Instructor; NYU/Stern
- Fall 2002: Information Technology and Electronic Commerce (MBA Core Class); Instructor; NYU/Stern
- Spring 2002: Information Technology and Electronic Commerce (MBA Core Class); Instructor; NYU/Stern
- Fall 2001: Information Technology and Electronic Commerce (MBA Core Class); Instructor; NYU/Stern
- Spring 2001: Information Technology and Electronic Commerce (MBA Core Class); Instructor; NYU/Stern
- Fall 2000: Information and Internet Technologies. Developed (w/ V.Vassalos) (MBA Class); Instructor; NYU/Stern
- Spring 2000: Technical Information Systems Research (Ph.D. Seminar); Instructor; NYU/Stern
- Spring 2000: Information Technology and Electronic Commerce (MBA Core Class); Instructor; NYU/Stern
- Spring 2000: Information and Internet Technologies. Developed (w/ V.Vassalos) (MBA Class); Instructor; NYU/Stern

PUBLICATIONS

Papers Submitted/Under Revision/In Preparation

- **2016: Who's Watching TV?**; Jessica Clark, Jean Francois Paiement, and Foster Provost; . Citations: 5
- **: On the Value of Data in Predictive Modeling**; Brian Dalessandro, Carlos Fernández, David Martens, Julie Moeyersoms, and Foster Provost;
- **: Classification Over Bipartite Graphs through Projection**; David Martens, Foster Provost, and Marija Stankova; . Citations: 16

Music

- **2018: Mean Reversion (Original Progressive Rock) Album**; Cliff Hurvich and Foster Provost; Kloster Art Music
- **2018: Blessing of My Birth (Original Progressive Rock) Single**; Cliff Hurvich and Foster Provost; Kloster Art Music
- **2018: Never Going to Work Again**; Foster Provost and Al Selvin; Kloster Art Music

Books

- **2013: Data Science for Business: What you need to know about Data Mining and Data Analytic Thinking**; Tom Fawcett and Foster Provost; O'Reilly Media. Citations: 3766
- **2001: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**; Foster Provost and Ramakrishnan Srikant; ACM Press

Invited White Paper

- **2017: Big Data, Data Science, and Civil Rights**; Solon Barocas, Elizabeth Bradley, Vasant Honavar, and Foster Provost; Computing Community Consortium of the Computing Research Association (CRA). Citations: 56

Journal Articles

- **2022: Explaining Data-Driven Decisions made by AI Systems: The Counterfactual Approach**; Carlos Fernández-Loría, Foster Provost, and Xintian Han; MIS Quarterly. Citations: 215
- **2022: A Comparison of Methods for Treatment Assignment with an Application to Playlist Generation**; Carlos Fernández-Loría, Foster Provost, Jesse Anderton, Benjamin Carterette, and Praveen Chandar; Information Systems Research. Citations: 34
- **2022: Causal Decision Making and Causal Effect Estimation Are Not the Same… and Why It Matters**; Carlos Fernández-Loría and Foster Provost; INFORMS Journal on Data Science. Citations: 110
- **2022: Causal Classification: Treatment Effect Estimation vs. Outcome Estimation**; Carlos Fernández and Foster Provost; Journal of Machine Learning Research. Citations: 74
- **2021: Classification over bipartite graphs through projection**; M. Stankova, S. Praet, D. Martens, and F. Provost; Machine Learning. Citations: 16
- **2020: A Benchmarking Study of Classification Techniques for Behavioral Data**; Sofie De Cnudde, David Martens , Theodoros Evgeniou, and Foster Provost; Int J Data Sci Analytics. Citations: 28
- **2020: Instance-level explanation algorithms SEDC, LIME, SHAP for behavioral and textual data: a counterfactual-oriented comparison**; Y. Ramon, D. Martens, F. Provost, and T. Evgeniou; Advances in Data Analysis and Classification
- **2019: Deep Learning on Big, Sparse, Behavioral Data**; Sofie De Cnudde, Yanou Ramon, David Martens, and Foster Provost; Big Data. Citations: 21
- **2019: Unsupervised Dimensionality Reduction vs. Supervised Regularization for Classification from Sparse Data**; Jessica Clark and Foster Provost; Data Mining and Knowledge Discovery
- **2018: Data-Driven Investment Strategies for Peer-to-Peer Lending**; Maxime Cohen, Daniel Guetta, Kevin Jiao, and Foster Provost; Big Data
- **2018: Wallenius Nai?ve Bayes**; Enric Junque de Fortuny, David Martens, and Foster Provost; Machine Learning. Citations: 2
- **2017: What Managers need to know about Big Data**; Jim Euchner and Foster Provost; Research Technology Management. Citations: 1
- **2017: In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions**; Zachery Anderson, Eva Ascarza, Peter Fader, Sunil Gupta, Bruce Hardie, Aurélie Lemmens, Barak Libai, Barak Libai, Scott Neslin, Oded Netzer, Foster Provost, and Rom Schrift; Customer Needs and Solutions. Citations: 393
- **2017: Privacy, Transparency and Control for Predictive Analytics from Massive Fine-Grained Personal Data**; Daizhuo Chen, Samuel Fraiberger, Robert Moakler, and Foster Provost; Big Data
- **2017: Cost-Effective Quality Assurance in Crowd Labeling**; Panagiotis Ipeirotis, Foster Provost, and Jing Wang; Information Systems Research. Citations: 95
- **2016: Mining Massive Fine-Grained Behavior Data to Improve Predictive Analytics**; Jessica Clark, Enric Junqué de Fortuny, David Martens, and Foster Provost; MIS Quarterly, Vol. 40 No. 4. Citations: 229
- **2015: Finding Mobile Consumers with a Privacy-Friendly Geo-Similarity Network**; David Martens, Alan Murray, and Foster Provost; Information Systems Research. Citations: 80
- **2015: Evaluating and Optimizing Online Advertising: Forget the Click, but there are Good Proxies**; Brian Dalessandro, Rod Hook, Claudia Perlich, and Foster Provost; Big Data. Citations: 76
- **2015: Beat the Machine: Challenging Humans to Find a Predictive Model's "Unknown Unknowns"**; Joshua Attenberg, Panos Ipeirotis, and Foster Provost; ACM Journal of Data and Information Quality. Citations: 244
- **2014: A Data Scientist's Guide to Start-Ups**; Ron Bekkerman, Oren Etzioni, Usama Fayyad, Claudia Perlich, Foster Provost, and Geoffrey Webb; Big Data. Citations: 1
- **2014: Machine Learning for Targeted Display Advertising: Transfer Learning in Action**; Brian Dalessandro, Claudia Perlich, Foster Provost, Troy Raeder, and Ori Stitelman; Machine Learning. Citations: 292
- **2014: Explaining Data-Driven Document Classifications**; David Martens and Foster Provost; MIS Quarterly. Citations: 463
- **2014: Repeated Labeling using Multiple Noisy Labelers**; Panagiotis Ipeirotis, Foster Provost, Victor Sheng, and Jing Wang; Data Mining and Knowledge Discovery. Citations: 249
- **2013: Predictive Modeling with Big Data: Is Bigger Really Better?**; Enric Junqué de Fortuny, David Martens, and Foster Provost; Big Data. Citations: 325
- **2013: Information in Digital, Economic and Social Networks**; Sinan Aral, Gal Oestreicher-Singer, Foster Provost, and Arun Sundararajan; Information Systems Research. Citations: 216
- **2013: Data Science and its Relationship to Big Data and Data-Driven Decision Making**; Tom Fawcett and Foster Provost; Big Data. Citations: 3027

- **2010: Inactive Learning?: Difficulties Employing Active Learning in Practice**; Josh Attenberg and Foster Provost; SIGKDD Explorations. Citations: 139
- **2009: Active Feature-Value Acquisition**; Prem Melville, Foster Provost, and Maytal Saar-Tsechansky; Management Science. Citations: 195
- **2007: Handling Missing Features when Applying Classification Models**; Foster Provost and Maytal Saar-Tsechansky; Journal of Machine Learning Research. Citations: 580
- **2007: Classification in Networked Data: A Toolkit and a Univariate Case Study**; Sofus Macskassy and Foster Provost; Journal of Machine Learning Research. Citations: 731
- **2007: Decision-Centric Active Learning of Binary-Outcome Models**; Foster Provost and Maytal Saar-Tsechansky; Information Systems Research. Citations: 81
- **2006: Network-Based Marketing: Identifying likely Adopters via Consumer Networks**; Shawndra Hill, Foster Provost, and Chris Volinsky; Statistical Science. Citations: 916
- **2006: Distribution-Based Aggregation for Relational Learning with Identifier Attributes**; Claudia Perlich and Foster Provost; Machine Learning. Citations: 152
- **2005: Intelligent Assistance for the Data Mining Process: An Ontology-Based Approach**; Abraham Bernstein, Shawndra Hill, and Foster Provost; IEEE Transactions on Knowledge and Data Engineering. Citations: 29
- **2004: Active Sampling for Class Probability Estimation and Ranking**; Foster Provost and Maytal Saar-Tsechansky; Machine Learning; 24. Citations: 281
- **2003: The Myth of the Double-Blind Review?: Author Identification Using Only Citations**; Shawndra Hill and Foster Provost; SIGKDD Explorations. Citations: 167
- **2003: Predicting Citation Rates for Physics Papers: Constructing Features for an Ordered Probit Model**; Sofus Macskassy, Claudia Perlich, and Foster Provost; SIGKDD Explorations. Citations: 4
- **2003: Learning when Training Data are Costly: The Effect of Class Distribution on Tree Induction** ; Foster Provost and Gary Weiss; Journal of Artificial Intelligence Research. Citations: 1348
- **2003: Tree Induction vs. Logistic Regression: A Learning-Curve Analysis**; Claudia Perlich, Foster Provost, and Jeffrey Simonoff; Journal of Machine Learning Research. Citations: 561
- **2003: Tree Induction for Probability-Based Ranking**; Pedro Domingos and Foster Provost; Machine Learning. Citations: 743
- **2001: Robust Classification for Imprecise Environments**; Tom Fawcett and Foster Provost; Machine Learning. Citations: 1941
- **2000: Discovering Interesting Patterns for Investment Decision Making with GLOWER — A Genetic Learner Overlaid with Entropy Reduction**; Dashin Chou, Vasant Dhar, and Foster Provost; Data Mining and Knowledge Discovery. Citations: 122
- **1999: A Survey of Methods for Scaling Up Inductive Algorithms**; Venkateswarlu Kolluri and Foster Provost; Data Mining and Knowledge Discovery. Citations: 374
- **1999: Problem Definition, Data Cleaning and Evaluation: A Classifier Learning Case Study**; Andrea Danyluk and Foster Provost; Informatica. Citations: 10
- **1997: Adaptive Fraud Detection**; Tom Fawcett and Foster Provost; Data Mining and Knowledge Discovery. Citations: 1487
- **1996: Scaling Up Machine Learning with Massive Parallelism**; John Aronis, Foster Provost, and Douglas Fisher (Editor); Machine Learning. Citations: 101
- **1995: Inductive Policy: The Pragmatics of Bias Selection**; Bruce Buchanan and Foster Provost; Machine Learning 20 (1/2):35-61. Citations: 156
- **1995: The Ten Most Common Plant Exposures Reported to Poison Information Centers in the United States**; Edward Krenzelok and Foster Provost; Journal of Natural Toxins
- **1992: A Distributed Algorithm for Embedding Trees in Hypercubes with Modifications for Run-Time Fault Tolerance**; Rami Melhem and Foster Provost; Journal of Parallel and Distributed Computing. Citations: 17
- **1990: A Real-Time Expert System for Trigger Logic Monitoring**; Scott Clearwater, Wilfred Cleland, Foster Provost, Eric Stern, and Zhenghai Zhang; Nuclear Instruments and Methods in Physics Research A293. Citations: 11
- **: Examining Dimensionality Reduction for Predictive Modeling**; Jessica Clark, and Foster Provost; Data Mining and Knowledge Discovery

Other Journal Articles

- **2002: Invited Comment on Bolton and Hand's "Statistical Fraud Detection"**; Richard Bolton, David Hand, and Foster Provost (Commentary); Statistical Science. Citations: 9
- **2001: Applications of Data Mining to Electronic Commerce**; Ron Kohavi and Foster Provost; Data Mining and Knowledge Discovery. Citations: 234
- **1998: On Applied Research in Machine Learning**; Ron Kohavi and Foster Provost; Machine Learning 30 (2/3) 1998. Citations: 326

Selected Conference Papers (Highly Selective Conferences)

- **2015: Measuring Causal Impact of Online Actions via Natural Experiments: Application to Display Advertising**; Daniel Hill, Alan Hubbard, Robert Moakler, Foster Provost, Kiril Tsemekhman, and Vadim Tsemekhman; Twenty-first ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2015). Citations: 42
- **2015: Iteratively Refining SVMs using Priors**; Theodoros Evgeniou, Enric Junque de Fortuny, David Martens, and Foster Provost; IEEE International Conference on Big Data (IEEE BigData 2015). Citations: 1
- **2014: Corporate Residence Fraud Detection**; Enric Junqué de Fortuny, David Martens, Bart Minnaert, Julie Moeyersoms, Foster Provost, and Marija Stankova; Twentieth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2014). Citations: 49
- **2014: Scalable Hands-Free Transfer Learning for Online Advertising**; Daizhuo Chen, Brian Dalessandro, Melinda Han Williams, Claudia Perlich, Foster Provost, and Troy Raeder; Twentieth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2014). Citations: 58
- **2013: Scalable Supervised Dimensionality Reduction using Clustering**; Brian Dalessandro, Claudia Perlich, Foster Provost, Troy Raeder, and Ori Stitelman; Nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2013). Citations: 19
- **2013: Using Co-Visitation Networks For Classifying Non-Intentional Traffic**; Brian Dalessandro, Rod Hook, Claudia Perlich, Foster Provost, Troy Raeder, and Ori Stitelman; Nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2013). Citations: 5
- **2012: Design Principles of Massive, Robust Prediction Systems**; Claudia Perlich, Brian Dalessandro, Foster Provost, Troy Raeder, and Ori Stitelman; Eighteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012). Citations: 60
- **2012: Bid Optimizing and Inventory Scoring in Targeted Online Advertising**; Claudia Perlich, Brian Dalessandro, Rod Hook, Foster Provost, Troy Raeder, and Ori Stitelman; Eighteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012).. Citations: 250
- **2011: Online Active Inference and Learning**; Josh Attenberg and Foster Provost; Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011).. Citations: 35
- **2010: A Unified Approach to Active Dual Supervision for Labeling Features and Examples**; Josh Attenberg, Prem Melville, and Foster Provost; European Conference on Machine Learning and Principles of Knowledge Discovery in Databases (ECML PKDD 2010). Citations: 78
- **2010: Why Label When You can Search? Alternatives to Active Learning for Applying Human Resources to Build Classification Models Under Extreme Class Imbalance**; Josh Attenberg and Foster Provost; Sixteenth ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD 2010). Citations: 120
- **2009: Audience Selection for On-line Brand Advertising: Privacy-Friendly Social Network Targeting**; Brian Dalessandro, Rod Hook, Alan Murray, Foster Provost, and Xiaohan Zhang; Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009). Citations: 207
- **2008: Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers**; Panagiotis Ipeirotis, Foster Provost, and Victor Sheng; Fourteenth ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2008. Citations: 1673
- **2007: Data Acquisition and Cost-Effective Predictive Modeling: Targeting Offers for Electronic Commerce**; Prem Melville, Foster Provost, and Maytal Saar-Tsechansky; Ninth International Conference on Electronic Commerce. Citations: 37
- **2005: An Expected Utility Approach to Active Feature-Value Acquisition**; Prem Melville, Raymond Mooney, Foster Provost, and Maytal Saar-Tsechansky; Fifth IEEE International Conference on Data Mining (ICDM-2005). Citations: 88
- **2005: ROC Confidence Bands: An Empirical Evaluation**; Sofus Macskassy, Foster Provost, and Saharon Rosset; 22nd International Conference on Machine Learning (ICML-2005); 27%. Citations: 103
- **2005: Suspicion Scoring Based on Guilt-by-Association, Collective Inference, and Focused Data Access**; Sofus Mcskassy and Foster Provost; International Conference on Intelligence Analysis (IA '05); 25%
- **2004: Active Feature-Value Acquisition for Classifier Induction**; Prem Melville, Raymond Mooney, Foster Provost, and Maytal Saar-Tsechansky; Fourth IEEE International Conference on Data Mining (ICDM2004). Citations: 127
- **2003: Aggregation-Based Feature Invention and Relational Concept Classes**; Claudia Perlich and Foster Provost; Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003); 13%. Citations: 160
- **2001: Intelligent Information Triage**; Vasant Dhar, Haym Hirsh, Sofus Macskassy, Foster Provost, and Ramesh Sankaranarayanan; 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2001); 23%. Citations: 53

- **2001: Active Learning for Class Probability Estimation and Ranking**; Foster Provost and Maytal Saar-Tsechansky; Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01); 24%. Citations: 2
- **1999: Activity Monitoring: Noticing Interesting Changes in Behavior**; Tom Fawcett and Foster Provost; Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99); 9%. Citations: 630
- **1999: Efficient Progressive Sampling**; David Jensen, Tim Oates, and Foster Provost; Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99); 9%. Citations: 572
- **1998: Robust Classification Systems for Imprecise Environments**; Tom Fawcett and Foster Provost; Fifteenth National Conference on Artificial Intelligence (AAAI-98); 30%. Citations: 174
- **1998: The Case against Accuracy Estimation for Comparing Classifiers**; Tom Fawcett, Ron Kohavi, and Foster Provost; Fifteenth Intl. Conference on Machine Learning (ICML-98); 31%
- **1997: Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions**; Tom Fawcett and Foster Provost; Third International Conference on Knowledge Discovery and Data Mining (KDD-97); 10%. Citations: 1245
- **1997: Scaling Up Inductive Algorithms: An Overview**; Venkateswarlu Kolluri and Foster Provost; Third International Conference on Knowledge Discovery and Data Mining (KDD-97); 41%. Citations: 40
- **1997: Increasing the Efficiency of Data Mining Algorithms with Breadth-First Marker Propagation**; John Aronis and Foster Provost; Third International Conference on Knowledge Discovery and Data Mining (KDD-97); 41%. Citations: 41
- **1996: Scaling Up: Distributed Machine Learning with Cooperation**; Daniel Hennessy and Foster Provost; Thirteenth National Conference on Artificial Intelligence (AAAI-96); 31%. Citations: 185
- **1996: Combining Data Mining and Machine Learning for Effective User Profiling**; Tom Fawcett and Foster Provost; Second International Conference on Knowledge Discovery and Data Mining (KDD96); 20%. Citations: 371
- **1996: Exploiting Background Knowledge in Automated Discovery**; John Aronis, Bruce Buchanan, and Foster Provost; Second International Conference on Knowledge Discovery and Data Mining (KDD-96); 33%. Citations: 50
- **1993: Iterative Weakening: Optimal and Near-Optimal Policies for the Selection of Search Bias**; Foster Provost; Eleventh National Conf. on Artificial Intelligence (AAAI-93); 25%. Citations: 21
- **1993: Small Disjuncts in Action: Learning to Diagnose Errors in the Telephone Network Local Loop**; Andrea Danyluk and Foster Provost; Tenth International Conference on Machine Learning (ICML93); 27%. Citations: 106
- **1992: Inductive Policy**; Bruce Buchanan and Foster Provost; Tenth National Conference on Artificial Intelligence (AAAI-92); 21%. Citations: 30

Additional Conference Papers

- **2014: Revealing Life Events from Inferred Customer Similarity: A Predictive Modeling Approach** ; Enric Junque de Fortuny, Tingting Nian, and Foster Provost; 24th Workshop On Information Technology and Systems (WITS 2014); 38%
- **2013: Hyperlocal: Inferring Location of IP Addresses in Real-Time Bid Requests for Mobile Ads**; Tina Eliassi-Rad, Long Le, Lauren Moores, and Foster Provost; 6th ACM SIGSPATIAL International Workshop on Location-based Social Networks (LBSN 2013). Citations: 7
- **2005: Suspicion Scoring of Networked Entities Based on Guilt-by-Association, Collective Inference, and Focused Data Access**; Sofus Macskassy and Foster Provost; Annual Conference of the North American Association for Computational Social and Organizational Science (NAACSOS), 2005. Citations: 60
- **2005: NetKit-SRL: A Network Learning Toolkit and its use for Classification of Networked Data**; Sofus Macskassy and Foster Provost; Annual Conference of the North American Association for Computational Social and Organizational Science (NAACSOS), 2005. Citations: 17
- **2004: Knowledge Discovery Using Concept-Class Taxonomies**; Bruce Buchanan, Venkateswarlu Kolluri, Douglas Metzler, and Foster Provost; AI 2004: Advances in Artificial Intelligence: 17th Australian Joint Conference on Artificial Intelligence. Springer-Verlag Heidelberg. Citations: 1
- **1994: Distributed Machine Learning: Scaling up with Coarse-Grained Parallelism**; Daniel Hennessy and Foster Provost; Second International Conference on Intelligent Systems for Molecular Biology (ISMB-94). Citations: 47
- **1992: ClimBS: Searching the Bias Space**; Foster Provost; Fourth International IEEE Conference on Tools with Artificial Intelligence(TAI-92); 36%. Citations: 3
- **1991: Embedding Rings in Hypercubes for Run-Time Fault Tolerance**; Rami Melhem and Foster Provost; Fourth ISMM/IASTED Intl. Conference on Parallel and Distributed Computing and Systems, 1991

- **1990: RL4: A Tool for Knowledge-Based Induction**; Scott Clearwater and Foster Provost; Second International IEEE Conference on Tools for Artificial Intelligence (TAI-90).; 40%. Citations: 143

Other Workshop/Symposium Papers

- **2018: Detecting Employee Misconduct and Malfeasance**; Josh Attenberg, Jennifer Chin, Chathra Hendahewa, Panagiotis Ipeirotis, Foster Provost, Abe Stanway, Bernando Suryanto, Bharath Vivekananda Swamy, and George Valkanas; MIS2: Misinformation and Misbehavior Mining on the Web, 2018, Marina Del Rey, CA, USA. Citations: 1
- **2016: Explaining Classification Models Built on High-Dimensional Sparse Data**; Brian d'Alessandro, David Martens, Julie Moeyersoms, and Foster Provost; ICML-2016 Workshop on Human Interpretability in Machine Learning (WHI 2016). Citations: 18
- **2016: Enhancing Transparency and Control when Drawing Data-Driven Inferences about Individuals**; Daizhuo Chen, Samuel Fraiberger, Robert Moakler, and Foster Provost; ICML-2016 Workshop on Human Interpretability in Machine Learning (WHI 2016). Citations: 68
- **2014: Pleasing the Advertising Oracle: Probabilistic Prediction from Sampled, Aggregated Ground Truth**; Brian Dalessandro, Claudia Perlich, Foster Provost, and Melinda Han Williams; Eighth International Workshop on Data Mining for Online Advertising (ADKDD 2014).. Citations: 6
- **2014: Causal Impact of Online Advertisements using Viewability as a Method of Treatment**; Robert Moakler and Foster Provost; Winter Conference on Business Intelligence
- **2012: Causally Motivated Attribution for Online Advertising**; Brian Dalessandro, Claudia Perlich, Foster Provost, and Ori Stitelman; ADKDD 2012. Citations: 203
- **2012: Estimating Audience Interest Distributions Based on Audience Web Behavior**; Foster Provost, Tsemekhman Kiril, and Xiaohan Zhang; Winter Business Intelligence Conference. Citations: 7
- **2012: Geo-Social Network Advertising**; David Martens, Alan Murray, and Foster Provost; Winter Business Intelligence Conference, 2012
- **2011: Beat the Machine: Challenging Workers to Find the Unknown Unknowns**; Josh Attenberg, Panagiotis Ipeirotis, and Foster Provost; Third Human Computation Workshop (HCOMP 2011). Citations: 244
- **2011: Estimating the Effect of Online Display Advertising on Browser Conversion**; Brian Dalessandro, Claudia Perlich, Foster Provost, and Ori Stitelman; ADKDD 2011. Citations: 41
- **2011: Managing Crowdsourcing Workers**; Panagiotis Ipeirotis, Foster Provost, and Jing Wang; Winter Business Intelligence Conference, 2011. Citations: 57
- **2010: Quality Management on Amazon Mechanical Turk**; Panagiotis Ipeirotis, Foster Provost, and Jing Wang; Second Human Computation Workshop (KDD-HCOMP 2010). Citations: 1416
- **2010: Active Inference and Learning for Classifying Streams**; Josh Attenberg and Foster Provost; ICML-2010 Workshop on Budgeted Learning. Citations: 4
- **2010: Guided Feature Labeling for Budget-Sensitive Learning Under Extreme Class Imbalance**; Josh Attenberg, Prem Melville, and Foster Provost; ICML-2010 Workshop on Budgeted Learning. Citations: 13
- **2007: Learning and Inference in Massive Social Networks**; Shawndra Hill, Foster Provost, and Chris Volinsky; The 5th International Workshop on Mining and Learning with Graphs, August 2007. Citations: 28
- **2007: Social Network Collaborative Filtering: Preliminary Results**; Anindya Ghose, Foster Provost, and Rong Zheng; The Sixth Workshop on e-Business (WeB 2007). Citations: 20
- **2006: A Brief Survey of Machine Learning Methods for Classification in Networked Data and an Application to Suspicion Scoring**; Sofus Macskassy and Foster Provost; ICML 2006 Workshop. Citations: 37
- **2005: Economical Active Feature-Value Acquisition through Expected Utility Estimation**; Prem Melville, Raymond Mooney, Foster Provost, and Maytal Saar-Tsechansky; KDD-05 Workshop on Utility-based Data Mining. Citations: 53
- **2005: Pointwise ROC Confidence Bounds: An Empirical Evaluation**; Sofus Macskassy, Foster Provost, and Saharon Rosset; ICML-2005 Workshop on ROC Analysis in Machine Learning (ROCML-2005). Citations: 21
- **2005: Toward a Justification of Meta-Learning: Is the No Free Lunch Theorem a Show-Stopper?**; Christophe Giraud-Carrier and Foster Provost; ICML-2005 Workshop on Meta-Learning. Citations: 72
- **2005: The Gift of Gab: Evidence TelE-Commerce Firms can Profit from Viral Marketing**; Shawndra Hill, Foster Provost, and Chris Volinsky; First Interdisciplinary Symposium between Information Systems, Statistics and Related Fields. Decision and Information Technologies Department, Robert H. Smith School of Business, University of Maryland
- **2004: Confidence Bands for ROC Curves: Methods and an Empirical Study**; Sofus Macskassy and Foster Provost; First Workshop on ROC Analysis in AI. Citations: 132
- **2003: A Simple Relational Classifier**; Sofus Macskassy and Foster Provost; Second Workshop on MultiRelational Data Mining (MRDM-2003) at KDD-2003. Citations: 396

- **2003: The Relational Vector-Space Model and Industry Classification**; Abraham Bernstein, Scott Clearwater, and Foster Provost; IJCAI-2003 Workshop on Learning Statistical Models from Relational Data. Citations: 64
- **2003: Relational Learning Problems and Simple Models**; Claudia Perlich, Sofus Macskassy, and Foster Provost; IJCAI-2003 Workshop on Learning Statistical Models from Relational Data.. Citations: 20
- **2003: Aggregation and Concept Complexity in Relational Learning**; Claudia Perlich and Foster Provost; IJCAI-2003 Workshop on Learning Statistical Models from Relational Data. Citations: 1
- **2002: Discovering Knowledge from Relational Data Extracted from Business News**; Abraham Bernstein, Scott Clearwater, Shawndra Hill, Claudia Perlich, and Foster Provost; KDD-2002 Workshop on Multi-Relational Data Mining. Citations: 50
- **2001: An Intelligent Assistant for the Knowledge Discovery Process**; Abraham Bernstein and Foster Provost; IJCAI-01 Workshop on Wrappers for Performance Enhancement in Knowledge Discovery in Databases, 2001. Citations: 80
- **2001: Information Triage using Prospective Criteria**; Vasant Dhar, Haym Hirsh, Sofus Macskassy, Foster Provost, and Ramesh Sankaranarayanan; Eighth International Conference on User Modeling (UM-2001), Workshop on Machine Learning, Information Retrieval and User Modeling. Citations: 12
- **1997: The WoRLD: Knowledge Discovery from Multiple Distributed Databases**; John Aronis, Bruce Buchanan, Venkateswarlu Kolluri, and Foster Provost; Florida Artificial Intelligence Research Symposium (FLAIRS-97). Citations: 75
- **1996: On the Value of Applied Research in Machine Learning**; Foster Provost, Tom Fawcett, Andrea Danyluk, and Pat Riddle; In Proceedings of the First International Workshop on Machine Learning in Practice (ICML 1996). Citations: 3
- **1995: Assessing Patient Referral Patterns to a Health Care Facility in Plant Exposure Patients using Computer Artificial Intelligence**; John Aronis, Bruce Buchanan, Timothy Jacobsen, Edward Krenzelok, and Foster Provost; European Assn. of Poison Centres and Clinical Toxicologists Scientific Meeting. May 18-20, 1995, Krakow, Poland
- **1995: Learning from Bad Data**; Andrea Danyluk and Foster Provost; ML-95 Workshop on Applying Machine Learning in Practice. Citations: 18
- **1995: Poinsettia (Euphorbia Pulcherrima) Exposures Have Good Outcomes… Just as we Thought** ; John Aronis, Bruce Buchanan, Timothy Jacobsen, Edward Krenzelok, and Foster Provost; European Association of Poison Centres and Clinical Toxicologists Scientific Meeting
- **1994: Efficiently Constructing Relational Features from Background Knowledge for Inductive Machine Learning**; John Aronis and Foster Provost; AAAI-94 Workshop on Knowledge Discovery in Databases - KDD Workshop 1994. Citations: 29
- **1994: Goal-Directed Inductive Learning: Trading off Accuracy for Reduced Error Cost**; Foster Provost; AAAI Spring Symposium on Goal-Directed Learning. Citations: 25
- **1993: Adaptive Expert Systems: Applying Machine Learning to NYNEX MAX**; Andrea Danyluk and Foster Provost; AAAI-93 Workshop
- **1992: A Baseline Taxonomy of Bias Adjustment Policies**; Foster Provost; Proceedings of the ML-92 Workshop on Biases in Learning

Book Chapters & Other Publications

- **2011: Selective Data Acquisition for Machine Learning**; Josh Attenberg, Prem Melville, Foster Provost, and Maytal Saar-Tsechansky; B. Krishnapuram, S. Yu, and B. Rao (Eds.), Cost-Sensitive Machine Learning. Chapman & Hall/CRC. Citations: 23
- **2005: Aggregation for Predictive Modeling with Relational Data**; Claudia Perlich and Foster Provost; Encyclopedia of Data Warehousing and Mining
- **2002: Scalability**; Venkateswarlu Kolluri and Foster Provost; W. Kloesgen and J. Zytkow (eds.), Handbook of Data Mining and Knowledge Discovery, Oxford University Press. Citations: 1
- **2002: Telecommunications Network Diagnosis**; Andrea Danyluk and Foster Provost; W. Kloesgen and J. Zytkow (eds.), Handbook of Data Mining and Knowledge Discovery, Oxford University Press. Citations: 2
- **2002: Fraud Detection**; Tom Fawcett and Foster Provost; W. Kloesgen and J. Zytkow (eds.), Handbook of Data Mining and Knowledge Discovery, Oxford University Press. Citations: 42
- **2000: Learning with Imbalanced Data Sets 101**; Foster Provost; AAAI-2000 Workshop on Learning with Imbalanced Data Sets, 2000.. Citations: 31
- **2000: Progressive Sampling**; David Jensen, Tim Oates, and Foster Provost; H. Liu and H. Motoda (eds.), Instance Selection and Construction, A Data Mining Perspective, Kluwer Academic Publishers. Citations: 8
- **1999: AI Approaches to Time-Series Problems**; Andrea Danyluk, Tom Fawcett, and Foster Provost; Workshop report in AI Magazine

- **1998: AI Approaches to Fraud Detection and Risk Management**; Tom Fawcett, Ira Haimowitz, Foster Provost, and Salvatore Stolfo; AI Magazine. Citations: 26
- **1992: Inductive Strengthening: The Effects of a Simple Heuristic for Restricting Hypothesis Space Search**; Bruce Buchanan and Foster Provost; K.P. Jantke (ed.), Analogical and Inductive Inference (Lecture Notes in Artificial Intelligence 642). Springer-Verlag. Citations: 10
- **1990: A Real-Time Expert System for Experimental High Energy/Nuclear Physics**; Scott Clearwater, Wilfred Cleland, Foster Provost, Eric Stern, and Zhenghai Zhang; D. Perrett-Gallix and W. Wojcik (eds.), New Computing Techniques in Physics Research. Paris: Centre National de la Recherche Scientific, 1990
- **1989: Distributed Fault Tolerant Embedding of Trees and Rings in Hypercubes**; Rami Melhem and Foster Provost; I. Koren (ed.) Defect and Fault Tolerance in VLSI systems Volume 1. New York, NY: Plenum Press

Other Working Papers

- **2011: Pseudo-Social Network Targeting from Consumer Transaction Data**; David Martens and Foster Provost; CeDER-11-05. Citations: 53
- **2008: Social Network Collaborative Filtering**; Foster Provost, Dennis Wilkinson, and Rong Zheng; CeDER- 8-08. Center for Digital Economy Research, Stern School of Business, New York University. Citations: 62
- **2000: Distributed Data Mining: Scaling Up and Beyond**; Foster Provost; H. Kargupta and P. Chan (eds.), Advances in Distributed Data Mining, San Francisco, CA. Citations: 80
- **2000: Well-Trained PETs: Improving Probability Estimation Trees**; Pedro Domingos and Foster Provost; CeDER Working Paper #IS-00-04. Citations: 226
- **1999: Rule-Space Search for Knowledge-Based Discovery**; John Aronis, Bruce Buchanan, and Foster Provost; CIIO Working Paper #IS 99-012. Citations: 24

Selected Tutorials

- **2016: Predictive Modeling with Social Networks**; Foster Provost; Conference on Complex Networks: from theory to interdisciplinary applications. Citations: 7
- **2009: Predictive Modeling with Social Networks**; Jennifer Neville and Foster Provost; Third International AAAI Conference on Weblogs and Social Media. Citations: 7
- **2008: Predictive Modeling with Social Networks**; Jennifer Neville and Foster Provost; Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Citations: 7
- **2008: Social Network Mining: A Tutorial on Inference and Learning with Social Network Data**; Jennifer Neville and Foster Provost; Twenty-Third AAAI Conf. on Artificial Intelligence
- **2007: Inference and Learning with Networked Data**; Foster Provost; NATO Advanced Study Institute on Mining Massive Data Sets for Security
- **2007: Modeling Complex Networks for Electronic Commerce**; Foster Provost and Arun Sundararajan; ACM Conference on Electronic Commerce. Citations: 2
- **1999: Evaluating Machine Learning, Knowledge Discovery, and Data Mining**; David Jensen and Foster Provost; KDD-98, at AAAI-99 and at IJCAI-99

SERVICE

- **University Service**
  - Director (interim) of NYU's Center for Data Science
  - NYU Committee on Intellectual Property Policy
  - Steering Committee for NYU Moore Sloan Data Science Environment Initiative-- $37.8M program with Berkeley and U.Wash.
  - NYU Venture Fund Investment Review Board
  - Skirball Center Academic Advisory Committee
  - NYU MSDS Curriculum Committee
  - University Working Group for 2014 Middle States Self-Study
  - University Working Group on Data Science Center and Data Science degree programs
- **School/Department Service**
  - Director: Fubon Center, Data Analytics and AI
  - Chair of Faculty Recruiting, IOMS/IS
  - Faculty Liaison/Advisor to the Stern Business Analytics Clubs

- co-designed and organized analytics competitions, panels, and other events. Co-designed & Hosted MSBA Business Analytics Symposium, “Competing in the Connected Economy: Applications & Implications”
- Design & Host Stern Biz Analytics Club event, “Whiskey Analytics (with tasting)”
- Stern MSBA curriculum committee
- IOMS/IS curriculum committee
- Co-organized the IOMS Dept. 2 Foster Provost, Ph.D. Research Fest
- IS Group faculty recruiting committee
- Statistics Group faculty recruiting committee
- Deans’ Faculty Advisory Committee
- IS Group faculty recruiting committee
- Stern Ph.D. Program Oversight Committee
- Stern Ph.D. Program Oversight Committee
- Cochair of faculty recruiting committee for IS Group
- Chair of faculty recruiting for IS Group
- Ph.D. recruiting committee for IS group
- Led the design of proposal for school-wide incentive policy for faculty procurement of external funding
- MBA Portfolio Committee
- U.G. Dean’s Faculty Advisory Committee

- **Editor-in-Chief**
  - Machine Learning journal
- **Founding Board Member (elected)**
  - International Machine Learning Society
- **Program Chair/General Chair/Main Organizer**
  - IEEE DSAA 2018 - The 5th IEEE International Conference on Data Science and Advanced Analytics. (General CoChair – with Francesco Bonchi)
  - The 1st, 2nd, 3rd, 4th, 5th and 6th Workshops on Information in Networks (WIN), (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, 3rd, 4th, 5th and 6th Workshops on Information in Networks (WIN), (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
  - Machine Learning journal
  - Journal of Machine Learning Research
  - Journal of Artificial Intelligence Research
- **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).