

## Rayid Ghani

rayidghani@gmail.com

<http://www.rayidghani.com>

Center for Data Science & Public Policy

Harris School of Public Policy & Computation Institute

University of Chicago

---

### Summary

I'm interested in solving and training students, analysts, and executives to solve large-scale, high impact, social and public policy problems in areas such as education, healthcare, economic development, public safety, criminal justice, and sustainability using data-driven and evidence-based methods from machine learning, data mining, data science, and social sciences.

- Director, Center for Data Science and Public Policy, University of Chicago
- Research Director & Senior Fellow, University of Chicago Computation Institute & Harris School of Public Policy.
- Founder & Director, Data Science for Social Good Fellowship Program, University of Chicago
- Co-founder of Edgeflip, a social network analytics driven startup focused on helping social causes.
- Chief Scientist for the Obama 2012 Campaign focusing on Analytics, Technology, and Data.
- Named “Young Global Leader” by the World Economic Forum in 2014
- 16 years of R&D experience in Data Analytics / Machine Learning / Data Science across politics, retail, healthcare, manufacturing, intelligence, and financial services industries.
- Led the Machine Learning & Data Mining Research group at Accenture Labs.
- Over 70 academic publications with more than 2000 citations, more than 15 patents awarded, and numerous media interviews and mentions in newspapers, magazines, TV, and books.
- Regular organizer and contributor to several international academic & industry conferences and workshops.
- Experience working with universities, government agencies, foundations, startups, and vendors on R&D and consulting projects.

### Research Expertise

Machine Learning, Data Science, Data Mining, Text Mining, Natural Language Processing, Information Retrieval, Social Media, Active Learning, Semi-Supervised Learning.

### Industry Expertise

Public Policy, Social Good, Non-Profits, Politics, Education, Public Health, Criminal Justice, Economic Development, Public Safety, Healthcare, Government, Retail, CPG, Financial Services, Manufacturing, Security.

### Work Experience

**Research Director & Senior Fellow**, Computation Institute, University of Chicago April 2013- Present

**Senior Fellow**, Harris School of Public Policy, University of Chicago

**Director and Founder**, Center for Data Science and Public Policy, University of Chicago

**Director and Founder**, Data Science for Social Good Fellowship, University of Chicago

Teaching at the Harris School of Public Policy and leading the Center for Data Science and Public Policy and our efforts on working with government agencies and non-profits using data science. Projects include:

- Creating and Running the Eric & Wendy Schmidt Data Science for Social Good Fellowship that brings 40-50 data science fellows for 12 weeks to work on social good projects in collaboration with government agencies and non-profits.
- Applied Data Analytics for Government training program for government agencies and non-profits.

- Working with police departments (Charlotte Mecklenburg, Nashville, Pittsburgh, Knoxville, San Francisco among others) to build data-driven Early Intervention Systems to reduce adverse police incidents with the public.
- Working with Chicago Department of Public Health to reduce Lead Poisoning through early, preventative home inspections.
- Working with Johnson County, KS and Salt Lake County, UT as part of the Data-Driven Justice Initiative to reduce incarceration.
- Work with school districts (Mesa AZ, Montgomery County MD, Vancouver WA, Cabarrus NC, Wake County NC, Tulsa OK, Muskingum Valley OH) to improve educational outcomes and tackle college under-matching and high school graduation.
- Work with City of Cincinnati to target home inspections to reduce urban blight.
- Teaching graduate students in social sciences and public policy: Machine Learning for Public Policy, Data Science and Policy Lab, Data Analytics for Campaigns.

**Co-Founder**  
Edgeflip

Jan 2013 – May 2015  
Chicago, IL

Co-founder of Edgeflip, an analytics-based startup focusing on helping social causes turn their influence into action. Edgeflip is building technology, data, and analytics products for non-profits, charities, and other social good organizations to better use social networks to do fundraising, volunteer recruiting and engagement, content sharing and advocacy.

**Chief Scientist**

Obama for America (Obama 2012 Election Campaign)

July 2011 – Dec 2012  
Chicago, IL

Chief (Data) Scientist for the Obama 2012 Campaign focusing on Analytics, Technology, and Data for improving different functions of the campaign including fundraising, volunteer, and voter targeting and mobilization. Led Data Science efforts at the campaign, especially focused on digital channels to help re-elect President Obama.

**Senior Research Scientist and Director - Analytics Research Group**

Accenture Technology Labs

July 2001 – July 2011  
Chicago, IL

Led a team of researchers focused on Applied Research in Machine Learning & Data Mining at Accenture Labs. Responsibilities included:

- Defining the Analytics R&D Agenda and Vision for Accenture Labs
- Defining and executing R&D projects (both internally and for clients) with a team of researchers and developers
- Identifying new business opportunities for Accenture that can be enabled by Analytics
- Conducting workshops with clients on Analytics and consulting on trends and emerging technologies in Analytics

Research focused on machine learning, text mining, and related areas motivated by business problems with high research and business impact. Projects ranged from internal ones, as well as with academic partners and clients such as Pfizer, eBay, USAA, Boeing, Microsoft, Best Buy, Pacificare, Kaiser, Wellpoint, GE, Ford, AT&T, US Govt Agencies.

Sampling of projects:

- Text Analytics for Email Monitoring and Compliance
- Social Media Analytics – Adaptive Sentiment Detection and Content Analysis
- Predictive Analytics for Predicting Errors in Insurance Claims Processing
- Data Anonymization & Text Redaction
- Text Analytics and Personalization for Enterprise Search & Knowledge Management

- Consumer Behavior Prediction & Personalized Advertising
- Text Analytics for Product Attribute Discovery & Extraction - Product Information Management
- Price Prediction in Online Markets

## Education

**Carnegie Mellon University**, Pittsburgh, PA,  
Machine Learning Department  
M.S. Machine Learning, 2001  
Advisor: Tom Mitchell

**University of the South**, Sewanee TN  
B.S (with Honors) Summa Cum Laude, 1999  
Majors: Computer Science, Mathematics

## Publications

### Books and Journals

Big Data and Social Science: Theory and Practical Approaches. Ian Foster, Rayid Ghani, Ron Jarmin and Frauke Kreuter, Julia Lane. CRC Press. 2016.

Data Mining for Business Applications. Carlos Soares, Rayid Ghani. IOS Press. 2010.

IEEE Transaction on Big Data. Special Issue on Urban Computing. Yu Zheng, Claudio Silva, Rayid Ghani, Cecilia Mascolo (Guest Editors). 2016.

Big Data Journal - Special Issue on Data for Good. Charlie Catlett and Rayid Ghani (Guest Editors). March 2015

### Edited Proceedings for Conferences and Workshops

The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2014. Sofus Macskassy, Claudia Perlich, Jure Leskovec, Wei Wang, Rayid Ghani, Prem Melville (Eds.). New York, NY, USA. ACM 2014.

Proceedings of the KDD Workshop on Data Science for Social Good. Arindam Banerjee, Lise Getoor, Rayid Ghani, Claire Montelioni, Matt Rattigan (Eds.). KDD Workshop 2014.

The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013. Inderjit S. Dhillon, Yehuda Koren, Rayid Ghani, Ted E. Senator, Paul Bradley, Rajesh Parekh, Jingrui He, Robert L. Grossman, Ramasamy Uthurusamy (Eds.). Chicago, IL, USA, August 11-14, 2013. ACM 2013, ISBN 978-1-4503-2174-7.

Proceedings of KDD Workshop on Data Mining for Business Applications. Rayid Ghani , Carlos Soares, Editors. *KDD Workshop 2008*.

Proceedings of KDD Workshop on Data Mining for Business Applications. Rayid Ghani , Carlos Soares, Editors. *KDD Workshop 2006*.

Learning from Partially Classified Data. M. Amini, O. Chapelle, R. Ghani, Editors. *Proceedings of ICML Workshop on Learning from Partially Classified Data* (2005).

Ghani, R., R. Jones, and C. Rosenberg. "The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining." *Workshop*. Vol. 951. No. 03. 2003.

### Book Chapters

Machine Learning. Rayid Ghani and Malte Schierholz. In Big Data and Social Science Research: Theory and Practical Approaches. Taylor Francis, 2016.

Machine Learning and Semantic Technologies for Enterprise Knowledge Management. Rayid Ghani and Divna Djordjevic. Book Chapter. Context and Semantics in Knowledge Management. Springer, 2011.

Incorporating Social Media Data into Customer Analytics. Rayid Ghani and Sarah Bentley. Book Chapter. Social Media Management Handbook. (2010). Wiley Press.

Data Mining for Consumer Modeling and Personalized Promotions. Rayid Ghani, Chad Cumby, Andrew Fano, and Marko Krema. Book Chapter - Data Mining Methods and Applications. 2008

Extracting and using Attribute-Value pairs from product descriptions on the web. Katharina Probst, Rayid Ghani, Yan Liu, Marko Krema, and Andrew Fano. Book chapter – Web Mining. 2007

Mining the Web to Add Semantics to Retail Data Mining. Rayid Ghani. Invited Chapter. “Web Mining: From Web to Semantic Web”. Springer Lecture Notes in Artificial Intelligence, Vol. 3209. Berendt, B.; Hotho, A.; Mladenic, D.; van Someren, M.; Spiliopoulou, M.; Stumme, G. (Eds.) 2004.

### **Peer-Reviewed Journal, Conference and Workshop Papers**

Machine Learning for Social Services: A case study of prenatal case management in Illinois. Ian Pan, Laura B. Nolan, Rashida R. Brown, Romana Khan, Paul van der Boor, Daniel G. Harris, Rayid Ghani. *American Journal of Public Health*. Forthcoming, 2017.

Early Intervention Systems - Predicting Adverse Interactions Between Police and the Public. Jennifer Helsby, Samuel Carton, Kenneth Joseph, Ayesha Mahmud, Youngsoo Park, Andrea Navarrete, Klaus Ackermann, Joe Walsh, Lauren Haynes, Crystal Cody, Major Estella Patterson, Rayid Ghani. *Criminal Justice Policy Review*, 2017.

Detecting fraud, corruption, and collusion in international development contracts: The design of a proof-of-concept automated system. Grace, Emily, Ankit Rai, Elissa Redmiles, and Rayid Ghani. *In 2016 IEEE International Conference on Big Data (Big Data) pp. 1444-1453. IEEE, 2016.*

The Legislative Influence Detector: Finding Text Reuse in State Legislation. Burgess et al. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2016)*.

Identifying Police Officers at Risk of Adverse Events, Carton et al. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2016)*.

Designing Policy Recommendations to Reduce Home Abandonment in Mexico. Ackerman et al. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2016)*.

Identifying Earmarks in Congressional Bills. Khabisa et al. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2016)*.

A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes. Lakkaraju et al. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2015)*.

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2015)*.

Early Prediction of Code Blue Using Electronic Medical Records. Sriram Somanchi, Samrachana Adhikari, Allen Lin, Elena Eneva, and Rayid Ghani. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2015)*.

Who, When, and Why: A Machine Learning Approach to Prioritizing Students at Risk of not Graduating High School on Time. Everaldo Aguiar, Himabindu Lakkaraju, Nasir Bhanpuri, David Miller, Ben Yuhas, Kecia Addison, Shihching Liu, Marilyn Powell, and Rayid Ghani. *5th International Learning Analytics and Knowledge (LAK) Conference 2015*.

Empirical Comparison of Active Learning Strategies for Handling Temporal Drift. Mohit Kumar, Mohak Shah, Rayid Ghani, Zubin Abraham. *KDD 2015 Workshop on Interactive Data Exploration and Analytics*.

Early Code Blue Prediction Using Patient Medical Records. Sriram Somanchi, Samrachana Adhikari, Allen Lin,

Elena Eneva, and Rayid Ghani. *Workshop on Machine Learning for Clinical Data Analysis and Healthcare – held with NIPS 2013*.

Online Active Learning with Imbalanced Classes. Zahra Ferdowsi, Rayid Ghani, Raffaella Settini. *IEEE International Conference on Data Mining (ICDM 2013)*.

Top-10 Data Mining Case Studies. Gabor Melli et al. *International Journal of Information Technology & Decision Making Vol 11 issue 02*. 2012.

Interactive Learning for Efficiently Detecting Errors in Insurance Claims. Rayid Ghani and Mohit Kumar. *Proceedings of the Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2011)*.

A Machine Learning Based System for Semi-Automatically Redacting Documents. Chad Cumby and Rayid Ghani. *Proceedings of the 23rd Annual Conference on Innovative Applications of Artificial Intelligence (IAAI) 2011*.

Framework for interactive classification problems. Mohit Kumar, Rayid Ghani, Mohak Shah, Jaime Carbonell, Alex Rudnicky. *ICML Workshop on Combining Learning Strategies to Reduce Label Cost – held with ICML 2011*.

An Online Strategy for Safe Active Learning. Zahra Ferdowsi, Rayid Ghani, Mohit Kumar. *ICML Workshop on Combining Learning Strategies to Reduce Label Cost – held with ICML 2011*

Testing Software in an Age Of Data Privacy: A Balancing Act. Kunal Taneja, Mark Grechanik, Rayid Ghani and Tao Xie. *Joint meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE 2011)*.

Graphics Classification for Enterprise Knowledge Management. Divna Djordjevic, Rayid Ghani. *Workshop on Application of Data Mining and Modelling in Government and Industry, held with IEEE International Conference on Data Mining (ICDM)*. 2010

Process-centric Enterprise workspace based in Semantic Wiki. Divna Djordjevic, Rayid Ghani, Duncan Fullarton. *International Conference on Knowledge Management and Information Sharing (KMIS)*. 2010.

Inference Control to Protect Sensitive Information in Text Documents. Chad Cumby, Rayid Ghani. *ACM SIGKDD Workshop on Intelligence and Security Informatics held with ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2010)*. 2010

Data Mining to Predict and Prevent Errors in Healthcare Claims Processing. Mohit Kumar, Rayid Ghani, and Zhu-Song Mei. *Proceedings of the Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2010)*.

Online Cost-Sensitive Learning for Efficient Interactive Classification. Rayid Ghani and Mohit Kumar. *Budgeted Learning Workshop at the 27 th International Conference on Machine Learning ICML 2010*.

Toward Optimal Ordering of Prediction Tasks. Yiming Yang, Abhimanyu Lad, Rayid Ghani, Bryan Kiesel. *SLAM International Conference on Data Mining (SDM 2009)*.

Graph Structure Learning for Task Ordering. Y. Yang, H. Shu, B. Kiesel, C. Cumby, R. Ghani, K. Probst. *ICEIS 2009*

Improving Knowledge Worker Productivity - the Active integrated approach  
P. Warren, N. Kings, I. Thurlow, J. Davies, T. Buerger, E. Simperl, C. Ruiz, J. M. Gomez-Perez, V. Ermolayev, R. Ghani, M. Tilly, T. Bösser, A. Imtiaz. 2009, *BT Technology Journal* (2009)

ACTIVE - Enabling the Knowledge-Powered Enterprise: Semantic Technology for Knowledge Worker Productivity. Warren, P., Thurlow, I., Ghani, R., Probst, K., Jentzsch, E., Ermolayev, V.. In *Proc 2<sup>nd</sup> European Semantic Technology Conference*, Vienna, Austria, Sep. 29 - Oct. 3, 2008

Maximizing Privacy Under Data Distortion Constraints in Noise Perturbation Methods. Yaron Rachlin, Katharina Probst, Rayid Ghani. *KDD Workshop on Privacy 2008*

Trade-offs in the Use of Bayesian Filtering for Sensor Fusion. Anatole Gershman, Rayid Ghani, Damian Roqueiro,

and Gang Wei. *International Workshop on Knowledge Discovery from Sensor Data (Sensor-KDD'07) –held with ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2007)*.

Towards Interactive Active Learning in MultiView Feature Sets for Information Extraction. Katharina Probst, Rayid Ghani. *European Conference on Machine Learning (ECML/PKDD 2007)*.

Learning to Extract Attributes from Product Descriptions. Katharina Probst, Rayid Ghani, Yan Liu, Marko Krema, and Andrew Fano. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2007)*.

Data Mining for Business Applications. Rayid Ghani, Carlos Soares. *SIGKDD Explorations 2006 Vol 8, Issue 2(2006)*.

Semi-Supervised Learning to Extract Attribute-Value pairs from Product Descriptions on the Web. Katharina Probst, Rayid Ghani, Yan Liu, Marko Krema, and Andrew Fano. *Workshop on Web Mining - - held with the European Conference on Machine Learning (ECML/PKDD 2006)*.

Text Mining to Extract Product Attributes. Rayid Ghani, Katharina Probst, Yan Liu, Marko Krema, and Andrew Fano. *SIGKDD Explorations (2006)*.

Using Bayesian Reasoning From Sensor Network for Indoor Surveillance. Valery Petrushin, Gang Wei, Rayid Ghani and Anatole Gershman. *Workshop on Pervasive Technology Applied: Real-World Experiences with RFID and Sensor Networks – held with*

Learning Individual Consumer Models for Personalized Promotions: A Data Mining Case Study. Chad Cumby, Andrew Fano, Rayid Ghani, and Marko Krema. *Workshop on Data Mining for Business - held with the European Conference on Machine Learning (ECML/PKDD 2005)*.

Price Prediction and Insurance for Online Auctions. Rayid Ghani. *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2005)*.

Multiple Sensor Integration for Indoor Surveillance. Valery Petrushin, Gang Wei, Rayid Ghani and Anatole Gershman. *Multimedia Data Mining Workshop – held with 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2005)*

A Bayesian Framework for Robust Reasoning from Sensor Networks. Valery Petrushin, Rayid Ghani and Anatole Gershman. *2005 AAAI Spring Symposium on AI Technologies for Homeland Security*

Building Intelligent Shopping Assistants using Individual Consumer Models. Chad Cumby, Andrew Fano, Rayid Ghani and Marko Krema. *Proceedings of the 2005 International Conference on Intelligent User Interfaces (IUI 2005)*

Predicting the End-Prices of Online Auctions. Rayid Ghani and Hillery Simmons. *Workshop on Data Mining & Adaptive Modeling Methods for Economics & Management - held with the European Conference on Machine Learning (ECML/PKDD 2004)*.

Predicting Customer Grocery Shopping Lists from Point-of-Sale Purchase Data. Chad Cumby, Andrew Fano, Rayid Ghani, and Marko Krema. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*.

Active learning for information extraction with multiple view feature sets. Rosie Jones, Rayid Ghani, Tom Mitchell and Ellen Riloff. *Workshop on Adaptive Text Extraction and Mining at European Conference on Machine Learning (ECML/PKDD 2003)*.

Building Minority Language Corpora by Learning to Generate Web Search Queries. Rayid Ghani, Rosie Jones and Dunja Mladenic. *Journal of Knowledge and Information Systems (KAIS), 2003*.

Using Text Mining to Infer Semantic Attributes for Retail Data Mining. Rayid Ghani, Andrew Fano. *IEEE International Conference on Data Mining (ICDM 2002)*

Building Recommender Systems using a Knowledge Base of Product Semantics. Rayid Ghani, Andrew Fano. *Workshop on Recommender Systems and Personalization in Ecommerce at 2nd International Conference on Adaptive Hypermedia and*

*Adaptive Web Based Systems* (2002)

Combining Labeled and Unlabeled for Multiclass Text Classification. Rayid Ghani. *Proceedings of the 19<sup>th</sup> International Conference on Machine Learning (ICML 2002)*.

A Comparison of Efficacy of Bootstrapping Algorithms for Information Extraction. Rayid Ghani and Rosie Jones. *Proceedings on the Workshop on Linguistic Knowledge Acquisition at the Linguistic Resources and Evaluation Conference (LREC 2002)*.

A Study of Approaches for Hypertext Categorization. Yiming Yang, Sean Slattery and Rayid Ghani. *Journal of Intelligent Information Systems - Special Issue on Automatic Text Categorization* (2002).

Hypertext Categorization using Hyperlink Patterns and Meta Data. Rayid Ghani, Sean Slattery, and Yiming Yang. *Proceedings of the 18th International Conference on Machine Learning (ICML 2001)*.

Using Error-Correcting Codes for Efficient Text Classification with a Large Number of Categories. Rayid Ghani. *Masters Thesis. Carnegie Mellon University* (2001)

Combining Labeled and Unlabeled data for Text Classification with a Large Number of Categories. Rayid Ghani. *Proceedings of the First IEEE Conference on Data Mining (ICDM 2001)*

Using Error-Correcting Codes and Co-Training for Text Classification with a Large Number of Categories. Rayid Ghani. *Workshop on Text Mining at the First IEEE Conference on Data Mining* (2001)

Using the Web to Create Minority Language Corpora. Rayid Ghani, Rosie Jones, and Dunja Mladenic. *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM 2001)*.

Online Learning for Query Generation: Finding Documents Matching a Minority Concept on the Web. Rayid Ghani, Rosie Jones, and Dunja Mladenic. *Proceedings of the First International Conference on Web Intelligence* (2001).

Automatic Web Search Query Generation to Create Minority Language Corpora.. Rayid Ghani, Rosie Jones, and Dunja Mladenic. *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001)*.

Using Error-Correcting Codes for Text Classification. Rayid Ghani. *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*.

Analyzing the Effectiveness and Applicability of Co-Training. Kamal Nigam & Rayid Ghani. *Proceedings of the Ninth International Conference on Information and Knowledge Management (CIKM 2000)*.

Understanding the Behavior of Co-Training. Kamal Nigam & Rayid Ghani. *Proceedings of the Workshop on Text Mining at the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2000)*.

Learning a Monolingual Language Model from a Multilingual Text Database. Rayid Ghani & Rosie Jones. *Proceedings of the Ninth International Conference on Information and Knowledge Management (CIKM 2000)*.

Automatically Building a Corpus for a Minority Language from the Web. Rosie Jones & Rayid Ghani. *Proceedings of the Student Workshop at the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)*.

Data Mining on Symbolic Knowledge Extracted from the Web. Rayid Ghani, Rosie Jones, Dunja Mladenic, Kamal Nigam, Sean Slattery. *Proceedings of the Workshop on Text Mining at the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2000)*.

### **Conference Abstracts and Presentations**

Identifying Police Officers at Risk of Adverse Events. Sam Carton, Jen Helsby, Kenny Joseph, Ayesha Mahmud, Youngsoo Park, Joe Walsh, Lauren Haynes, and Rayid Ghani. Presented at Population Association of America Conference, Washington, DC, April 2016.

Healthier homes with predictive analytics to identify risks of lead poisoning. APHA 2015

## Media Mentions, Interviews, and Coverage

Featured in books such as *The Numerati* (Stephen Baker), *Predictive Analytics* (Eric Siegel), *The Master Algorithm* (Pedro Domingos), and *SuperCrunchers* (Ian Ayres)

Named as a *Young Global Leader* by the World Economic Forum in 2014

Named as one of the *50 Politicos to Watch*, by Politico Magazine in 2013.

Work featured in numerous media stories for publications such as *Time*, *New York Times*, *Slate*, *Business Week*, *NPR*, *Financial Times*, *Chicago Tribune*, *US News & World Report*, *CNN*, *Mother Jones*, *GovTech*, and *NBC*.

## Patents Granted

- Classification-based redaction in natural language text
- User modification of generative model for determining topics and sentiments
- Processing a reusable graphic in a document
- Claims analytics engine
- Sentiment classifiers based on feature extraction
- Promotion planning system
- System for individualized customer interaction
- Entity assessment and ranking
- Data anonymization based on guessing anonymity
- Preprocessing of Text
- Extraction of attributes and values from natural language documents
- Determination of a basis for a new domain model based on a plurality of learned models
- Preprocessing of text for Sentiment Classification
- Identification of attributes and values using multiple classifiers
- Automated classification algorithm comprising at least one input-invariant part
- Determination of a profile of an entity based on product descriptions
- Auction insurance system
- Auction result prediction

## Teaching

- Co-Created and Teaching “Applied Data Analytics for Public Policy” Executive Program in collaboration with Frauke Kreuter (University of Maryland) and Julia Lane (NYU). Spring 2017
- Machine Learning for Public Policy Lab. Winter 2017
- Created and Taught “Machine Learning for Public Policy” for graduate students in the Harris School of Public Policy as well as Social Sciences, University of Chicago. Spring 2015, Spring 2016, Spring 2017
- Created and Taught “Data Analytics for Campaigns” class for graduate students in the Harris School of Public Policy, University of Chicago. Winter 2015
- Big Data course designed for US Federal Agencies. Taught Machine Learning, Text Analysis, Network Analysis classes. Spring 2014, Fall 2014, Fall 2015
- Routinely teaching workshops and tutorials in Machine Learning and Data Science to students, analysts, and non-profits

## Student Advising



### **PhD Committees:**

- Sabina Tomkins (University of California, Santa Cruz)
- Zahra Ferdowsi (DePaul University). Graduated June 2015
- Mohit Kumar (Carnegie Mellon University), Graduated May 2014
- Runa Bhaumik. DePaul University. Graduated 2011.
- Yanzan Zhou. DePaul University. Graduated 2007
- Xin Jin. DePaul University. Graduated 2006.

### **Recent Grants**

- Schmidt Family Foundation – Data Science for Social Good Fellowship 2013 \$800,000
- Schmidt Family Foundation – Data Science for Social Good Fellowship 2014 \$1.5M
- Anonymous Foundation – Program Management Support \$260,000 – 2014-2015
- Schmidt Family Foundation – Data Science for Social Good Fellowship 2015 \$1.5M
- City of Cincinnati - Predictive Analytics to Increase Building Code Compliance 2015 \$60,000
- National Science Foundation – National Data Science Organizers Workshop 2015 \$32,000
- Schmidt Family Foundation – Data Science for Social Good Fellowship 2016 \$1.55M
- Trinity Hospital System (Gift) – Preventative Interventions to Improve Community Health 2016 \$99,000
- Laura and John Arnold Foundation – Urban Analytics Network 2016-2017 \$835,000
- Robert Wood Johnson Foundation – Data Across Sectors for Health 2016 \$60,000
- Chicago Department of Public Health (via CDC) – Predictive Analytics to Improve the HIV Spectrum of Care 2016 \$50,000
- Case Foundation – Impact Investing Network. \$50,000
- NRT-IGE: Information Infrastructure for Society: Integrating Data Science and Social Science in Graduate Education and Workforce Development. Co-PI. \$499,896
- Accenture (Gift) - \$100,000
- Center for Aids Research, Pilot Grant, \$50,000 2017

### **Boards**

- Technical Advisory Board, Hispanic Scholarship Fund (Non-profit)
- Board of Directors, ChangeLab Solutions (Non-profit focused on Public Health Law and Policy)
- WEF Value based Healthcare
- Advisory Board, Elastica (Network Security Startup)
- Alumni Board Member, Sewanee: The University of the South. 2010-2012
- Advisory Board, European Union funded Project on Semantically Enabled Knowledge Technologies (SEKT) 2004-2006

### **Invited Panels**

- Artificial Intelligence and Blockchain, organized by the Federal Trade Commission, Feb 2017
- Data Science for Policy and Social Good. Opening Event for Data Science for Social Good Europe, Cascais June 2017
- Winter Innovation Summit, Salt Lake City, UT. January 2017
- Australian AI conference, Hobart. December 2017
- Fairness, Accountability, and Transparency in Machine Learning Workshop 2016.
- The Future of the Criminal Justice System. White House Frontiers Conference, 2016.

- Data Science Education. Computing Research Association Meeting, 2016
- Artificial Intelligence for the Public Good. Workshop organized by CCC and Office of Science and Technology Policy, White House.
- Big Data and Advocacy Panel. Environmental Law & Policy Center, 2016.
- Big Data and AI: Sci-Fi vs Everyday Applications. SXSW Interactive Panel 2016
- Data to Drop Violence. PeaceTech Summit, 2016.
- Big Data Panel, at Independent Sector Conference, 2015
- Police Data Initiative Panel at International Association of Chiefs of Police Conference, IACP 2015
- #SocialFundamentals: How to use social media to effectively brand your nonprofit., OnBoard Conference, organized by the Social Enterprise Institute, March 2015.
- Does Social Good Justify Risking Personal Privacy, KDD 2014 Panel.
- City of Big Data Panel, University of Chicago, 2014
- Big Data Week: The Public Life of Big Data. Chicago, 2014.
- NIPS 2014 Workshop on Fairness, Accountability, and Transparency in Machine Learning.
- Data for Good, Strata, 2014.
- Pivot 2015
- Data and Campaigns, Strata 2013
- Voter Modeling Panel at Data Crunched Democracy Conference, University of Pennsylvania, 2013
- ND-GAIN Annual Meeting, 2013

## **(Recent) Keynote Talks at Conferences**

- Xad September 2016
- Liveramp March 2017
- PyData August 2016
- ESWC Summer School on Data Science August 2016
- Data and Policy Conference, Cambridge September 2016
- Australian AI conference December 2016
- Conference on Data Analysis - Department of Energy. Los Alamos National Labs. 2016
- Workshop on Data4Good: Machine Learning in Social Good Applications at the International Conference on Machine Learning, 2016.
- Chief Data Officer Executive Summit 2015
- Environmental Measurement Symposium, organized by the EPA, July 2015
- DRIVE Conference, March 2015.
- American Association for Artificial Intelligence/Innovative Applications of Artificial Intelligence (IAAI/AAAI) 2015 Keynote Talk
- KDD @ Bloomberg – Solve for Good, held in conjunction with KDD 2014.
- Machine Learning Symposium, New York Academy of Sciences, January 2014.
- Keynote at ECML-PKDD 2013
- Grace Hopper Celebration of Women in Computing Conference 2014
- Annual Summit on Strategic Communications 2013
- CeBIT Australia. May 2013.
- ACM SIGMETRICS Workshop. June 2013.
- Annual Conference for Australian Digital Marketing Association. August 2013.
- Large-Scale Online Learning Workshop, Cumberland Lodge, UK. September 2013
- Annual Future of Privacy Forum, Washington, DC.
- Predictive Analytics World, June 2013

- Predictive Analytics World, March 2013
- OpenCourseWare Consortium Global Conference 2014
- Internet of Education Conference, 2013
- Digital Marketing Collective Annual Meeting, 2013
- Digital Leaders Annual Lecture. UK Parliament, January 2013
- Big Data Business Forum, November 2012

## **Other Invited Talks**

- Midwest Machine Learning Symposium, June 2017
- University of the South, Sewanee April 2017
- Analytics workshop for Governments, Melbourne December 2016
- University College London - Public Policy and Computer Science Seminar
- White House Police Data Initiative. 2016.
- What Works Cities Summit. 2016.
- Harvard University March 2016
- CUSP NYU, January 2016
- UC Santa Cruz Research Day, November 2015
- Data Science Seminar Series, Berkeley Institute of Data Science, UC Berkeley. March 2015
- The Role of Machine Learning in the Social Sciences. Demography Workshop organized by the Population Research Center, University of Chicago. February 2015.
- Computational Social Science Seminar, University of Chicago.
- Independent Sector Conference - Embark 2015
- Beyond Big: The Analytically Powered Organization. Organized by ATKearney and Carnegie Mellon University, January 2014.
- Machine Learning and Data Analytics Symposium – MLDAS 2014, organized by Boeing and Qatar Computing Research Institute, 2014.
- MCON 2014
- How Relevance Can Get Your Brand Elected, National Retail Federation, 2014.
- Teradata Universe, Japan. 2014
- Teradata Universe, China. 2014
- CIO Workshop, Singapore, 2014.
- Clinton School of Public Service, University of Arkansas, Little Rock. 2014.
- Training Students to Extract Value from Big Data. National Academies, 2014
- Northwestern University, February 2013
- UCLA
- Techweek Chicago 2013
- RichRelevance Summit
- Khan Academy
- Maritz
- NY Data Meetup
- Pratham Annual Gala
- Innovation Congress, Rio Grande de Sul, Brazil, October 2014.
- University of California Santa Cruz Data Science Seminar
- KDD 2013 Industrial Practice Expo. August 2013.
- Teradata Universe, Denmark, 2013.
- Machine Learning Seminar - Carnegie Mellon University. February 2013.

- InfoSeminar, Stanford University. February 2013.
- Computer Science Department, UIC.
- School of Engineering, University of California, Santa Cruz
- AAAS Family Day
- Ignite
- Forum Analytics, 2013
- Do Good Data Conference 2013, 2014
- Accenture CIO Circle 2013
- iCrossing CMO Summit, 2013
- Twitter
- KDD Industrial Practice Expo Track 2011
- ACM Conference on Knowledge Discovery & Data Mining (Application Track) - KDD 2010
- Industry Track, European Conference on Machine Learning & Data Mining (ECML PKDD 2009)
- Workshop on Machine Learning for the Semantic Web at ESWC 2009
- Workshop on Data Mining for Advertising at KDD 2008
- Workshop on MultiView Algorithms at International Conference on Machine Learning (ICML 2005)
- Web Mining Workshop at European Conference on Machine Learning & Principles of Data Mining (ECML/PKDD 2003)
- Google Research
- Yahoo Research
- Toyota Technological Institute - Chicago
- DePaul University
- Notre Dame University

### **Events Organized:**

- NSF Workshop for National Data Science Organizers, 2015
- Workshop on Data Science for Social Good at KDD 2014, New York.
- Industry & Government Program Chair – KDD 2014
- Industry & Government Program Chair – KDD 2013
- Workshop on Machine Learning for Global Challenges – held with International Conference on Machine Learning, 2011
- Workshop on Data Mining for Medicine and Healthcare – held with 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011
- Workshop on Data Mining for Business Applications – held with 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008
- Workshop on Data Mining for Business Applications – held with 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006
- Workshop on Learning from Partially Classified Data – held with International Conference on Machine Learning, 2005
- Workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining – held with International Conference on Machine Learning, 2003
- Workshop on Operational Text Classification – held with 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003

### **Recent Professional Service Activities**

- Social Impact Chair, KDD 2016, KDD 2017
- Senior Program Committee Member, Applied Data Science Track, KDD 2016, 2017

- Guest Editor, Journal, Urban Computing Special Issue, 2016
- Guest Editor, Big Data Journal, Special Issue on Data for Social Good, 2015.
- Senior Program Committee Member, Research Track, KDD 2015
- Senior Program Committee Member, Industry Track, KDD 2015
- Panels Chair, IEEE ICDM 2015.
- Industry & Government Program Chair, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014 (KDD 2014)
- Industry & Government Program Chair, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013 (KDD 2013)
- Industry & Government Program Chair, IEEE Big Data Conference, 2013
- Media Chair, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012 (KDD 2012)
- Publicity Chair, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011 (KDD 2011)
- Machine Learning Track Chair, European Semantic Web Conference (ESWC 2011)
- Best Paper Award Committee & Senior Program Committee, KDD 2010
- Workshops Chair, European Conference on Machine Learning (ECML.PKDD 2009)
- Area Chair: ACM Conference on Knowledge and Information Management (CIKM 2008)
- Area Chair: European Conference on Machine Learning (ECML.PKDD 2007)
- NSF Review Panels, Spring 2004 and Spring 2005

### **Conference Program Committees:**

- ACM Conference on Knowledge Discovery & Data Mining (KDD)
- International Conference on Machine Learning (ICML)
- SIAM conference on Data Mining (SDM)
- IEEE International Conference on Data Mining (ICDM)
- ACM Conference on Research and Development in Information Retrieval (SIGIR).
- Annual Conference on Neural Information Processing Systems (NIPS)
- ACM Conference on Information and Knowledge Management (CIKM)
- National Conference on Artificial Intelligence (AAAI)
- European Conference on Artificial Intelligence (ECAI)
- International World Wide Web Conference (WWW)
- Open Course Ware Consortium (OCWC) Global Conference
- E-Commerce and Web Technologies-Conference
- International Conference on Data Warehousing and Knowledge Discovery - DaWaK
- IEEE International Conference on Tools with Artificial Intelligence (ICTAI)
- ICDM 2010 Workshop on Large-scale Analytics for Complex Instrumented Systems (LACIS 2010).
- Workshop on Large-scale Data Mining: Theory and Applications (LDMTA 2009)
- KDD workshop on Data Mining and Audience Intelligence for Online Advertising 2008-2011
- CIAO 2009-2010
- SEMSEARCH 2009
- Workshop on Information Retrieval and Advertising (at SIGIR 2009)
- Workshop on Intelligent Techniques for Web Personalization
- Workshop on Learning from Multiple Views – held with International Conference on Machine Learning, 2005
- Link Discovery Workshop (LinkKDD)
- Web Mining Workshop (WebKDD)

- Adaptive Text Extraction & Mining Workshop at ECML 2003
- Text Learning Workshop at the International Conference on Machine Learning (ICML 2002)
- Operational Text Classification Workshop at the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002)
- Text Mining Workshop at the IEEE Conference on Data Mining, 2001

**Reviewer for:**

- Machine Learning Journal (MLJ)
  - Journal of Artificial Intelligence Research (JAIR)
  - Journal of Machine Learning Research (JMLR)
  - Journal of Neurocomputing
  - IEEE Transaction on Knowledge and Data Engineering (TKDE)
  - ACM Transactions on Information Systems (TOIS)
  - European Conference on Artificial Intelligence (ECAI 2002)
  - International Conference on Machine Learning (ICML 2001).
-