Rayid Ghani

rayid@cmu.edu
http://www.rayidghani.com

Machine Learning Department and
Heinz College of Information Systems & Public Policy
Carnegie Mellon University

Work Experience

Distinguished Career Professor	September 2019 - Present

Joint Appointment in **the Machine Learning Department** and the
Heinz College of Information Systems and Public Policy
Carnegie Mellon University

Research Associate Professor	2018 – August 2019

Department of Computer Science**,** University of Chicago

Senior Fellow	2013 - August 2019

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

Research Director & Senior Fellow	2013 - 2018

Computation Institute**,** University of Chicago

Director and Founder	2013 - August 2019

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

Director and Founder	2013 - Present

Data Science for Social Good Fellowship

Co-Founder	2013 - 2015

Edgeflip
Co-founder of Edgeflip, a data science startup focusing on helping social causes turn their influence into action. Edgeflip built 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	2011 - 2012

Obama for America (Obama 2012 Election Campaign)

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	2001 - 2011
Director - Analytics Research Group	2006 - 2011

Accenture Labs

Led a team of researchers focused on Applied Research in Machine Learning & Data Science 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

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

Awards

Milbank Memorial Fund and AcademyHealth State and Local Innovation Prize – awarded to work with Chicago Department of Public Health on Lead Poisoning Prevention. 2018

IACP/Laura and John Arnold Foundation Leadership in Law Enforcement Research Award – awarded to work with Charlotte Mecklenburg Police Department. 2018

American Statistical Association Harry V. Roberts Statistical Advocate of the Year Award, 2017

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

Distinguished Young Alumni Award, Sewanee - University of the South, 2013

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

Media Coverage

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

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

Publications

Books

Big Data and Social Science: Theory and Practical Approaches. Ian Foster, Rayid Ghani, Ron Jarmin, Frauke Kreuter, and Julia Lane. CRC Press. 2016 (2nd Edition - 2020)

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

Edited Journals and Conference Proceedings

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

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 (updated 2nd edition, 2020)

Fairness in Machine Learning. Kit Rodolfa, Pedro Saleiro, Rayid Ghani. In Big Data and Social Science Research: Theory and Practical Approaches. Taylor Francis, 2020

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

Bandit Data-Driven Optimization. Zheyuan Ryan Shi, Zhiwei Steven Wu, Rayid Ghani, Fei Fang. AAAI-22: the 36th AAAI Conference on Artificial Intelligence (2022)

Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy. Rodolfa, K.T., Lamba, H. & Ghani, R. Nature Machine Intelligence 3, 896–904 (2021).

An Empirical Comparison of Bias Reduction Methods on Real-World Problems in High-Stakes Policy Settings. Hemank Lamba, Kit T Rodolfa, Rayid Ghani. ACM SIGKDD Explorations 2021.

Taking Our Medicine: Standardizing Data Science Education With Practice at the Core. Kit Rodolfa, Rayid Ghani. Harvard Data Science Review. https://doi.org/10.1162/99608f92.64bbf5af (2021)

Machine Learning Informed Decision-Making with Interpreted Model's Outputs: A Field Intervention. Leid Zejnilovic, Susana Lavado, Carlos Soares, Íñigo Martínez De Rituerto De Troya, Andrew Bell and Rayid Ghani. Academy of Management, 2021. Best Paper Runner-Up Award

Wilde, H., Chen, L., Nguyen, A., Kimpel, Z., Sidgwick, J., De Unanue, A., . . . Vollmer, S. (2021). A recommendation and risk classification system for connecting rough sleepers to essential outreach services. Data & Policy, 3, E2. doi:10.1017/dap.2020.23

Validation of a Machine Learning Model to Predict Childhood Lead Poisoning. Eric Potash, Rayid Ghani, Joe Walsh, Emile Jorgensen, Cortland Lohff, Nik Prachand, Raed Mansour. JAMA Network Open. 2020;3(9):e2012734. doi:10.1001/jamanetworkopen.2020.12734.

Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. Vollmer Sebastian, Mateen Bilal A, Bohner Gergo, Király Franz J, Ghani Rayid, Jonsson Pall et al. BMJ 2020; 368: l6927 (2020)

Predictive Fairness to Reduce Misdemeanor Recidivism Through Social Service Interventions. K. Rodolfa; E. Salomon; L. Haynes; I. Mendieta; J. Larson; R. Ghani. ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*), 2020.

Predictive Analytics for Retention in Care in an Urban HIV Clinic. Nature Scientific Reports 10, 6421 (2020). https://doi.org/10.1038/s41598-020-62729-x

Bandit Data-driven Optimization: AI for Social Good and Beyond. Zheyuan Ryan Shi, Zhiwei Steven Wu, Rayid Ghani and Fei Fang. ML for Economic Policy Workshop at NeurIPS 2020.

Mapping New Informal Settlements using Machine Learning and Time Series Satellite Images: An Application in the Venezuelan Migration Crisis (By: Isabelle Tingzon, Niccolo Dejito, Ren Avell Flores, Rodolfo De Guzman, Liliana Carvajal, Katerine Zapata Erazo, Ivan Enrique Contreras Cala, Jeffrey Villaveces, Daniela Rubio and Rayid Ghani. In Fragile Earth: Data Science for a Sustainable Planert – workshop at KDD 2020. Other versions at Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop at NeurIPS 2020. Longer version at 2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G).

An Experience-Centered Approach to Training Effective Data Scientists. Kit T. Rodolfa, Adolfo De Unanue, Matt Gee, and Rayid Ghani. Big Data. Dec 2019. 249-261

Change Through Data: A Data Analytics Training Program for Government Employees. Kreuter, F., Ghani, R., & Lane, J. Harvard Data Science Review, 1(2). 2019

Using machine learning to help vulnerable tenants in New York city. Teng Ye, Rebecca Johnson, Samantha Fu, Jerica Copeny, Bridgit Donnelly, Alex Freeman, Mirian Lima, Joe Walsh, and Rayid Ghani. In Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '19).

Deploying Machine Learning Models for Public Policy: A Framework. Klaus Ackermann, Joe Walsh, Adolfo De Unánue, Hareem Naveed, Andrea Navarrete Rivera, Sun-Joo Lee, Jason Bennett, Michael Defoe, Crystal Cody, Lauren Haynes and Rayid Ghani. 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2018).

Reducing Incarceration through Prioritized Interventions. Matthew J. Bauman, Kate Boxer, Tzu-Yun Lin, Erika Salomon, Hareem Naveed, Lauren Haynes, Joe Walsh, Jen Helsby, Steve Yoder, Robert Sullivan, Rayid Ghani. ACM SIGCAS Conference on Computing and Sustainable Societies, 2018.

Improving Government Response to Citizen Requests Online. Garren Gaut, Andrea Navarette, Laila Wahedi, Paul van der Boor, Adolfo de Unánue, Jorge Díaz, Eduardo Clark, Rayid Ghani. ACM SIGCAS Conference on Computing and Sustainable Societies, 2018.

Using Machine Learning to Assess the Risk of and Prevent Water Main Breaks. Avishek Kumar, Syed Ali Asad Rizvi, Benjamin Brooks, Ali Vanderveld, Kevin Hayes Wilson, Chad Kenney, Adria Finch, Andrew Maxwell, Sam Edelstein, Joe Zuckerbraun and Rayid Ghani. 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2018).

Visualizing Meta-Explanations in Early Intervention Systems for Police Departments [Poster]. Damon Crockett, Joe Walsh, Klaus Ackermann, Andrea Navarrete, Rayid Ghani. IEEE Information Visualization Conference 2017 (IEEE VIS 2017)

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, 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. Khabsa 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, Rafaella Settimi. 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 27th International Conference on Machine Learning ICML 2010.

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

Graph Structure Learning for Task Ordering.Y. Yang, H. Shu, B. Kisiel, 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 Technologiy 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 2nd 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 19th 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

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 Taught "Machine Learning in Practice" for graduate students from Machine Learning Department at Carnegie Mellon University, Fall 2020, Spring 2021, Fall 2021, Fall 2022
- Co-Created and Taught "Machine Learning for Public Policy Lab" for graduate students in Public Policy School at Carnegie Mellon University, Fall 2020, Spring 2021, Fall 2021, Fall 2022
- Co-Created and Taught a First-year Grand Challenge Seminar for School of Humanities and Social Science students on "Designing Better Human-AI Futures" Spring 2022
- 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, Spring 2018, Spring 2019
- Co-Created and Teaching "Applied Data Analytics for Public Policy" Executive Training Program in collaboration with Frauke Kreuter (U. of Maryland) and Julia Lane (NYU). 2017-Present
- Machine Learning for Public Policy Lab. Winter 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:

- Joachim Bauman, University of Zurich
- Lingwei Cheng – Heinz Paper Committee CMU
- Amanda Coston – Heinz Paper Committee CMU
- Pallavi Koppol – Computer Science Department (CMU)
- James Fiacco – Language Technologies Institute (CMU)
- Ryan Shi, ISR – Societal Computing PhD Program (CMU)
- Emily Black, Computer Science Department (CMU), Graduated May 2022
- Maryam Tabar, Penn State University
- Dylan Fitzpatrick (CMU), Graduated August 2020
- Teng Ye (University of Michigan, Graduated December 2021
- Victor Kristof (EPFL), Graduated June 2021
- Alejandro Noriega (MIT), Graduated
- 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.

Post-docs Supervised:
- Jen Helsby
- Eric Potash
- Erika Salomon
- Avishek Kumar
- Klaus Ackermann
- Arthi Ramachandaran
- Pedro Saleiro
- Kasun Amarasinghe
- Hemank Lamba

Recent Grants and Contracts

- 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
- Robert Wood Johnson Foundation – Data Across Sectors for Health 2016 $60,000
- Laura and John Arnold Foundation – Urban Analytics Network 2016-2017 $835,000
- Chicago Department of Public Health (via CDC) – Predictive Analytics to Improve the HIV Spectrum of Care 2017 $50.000
- Case Foundation – Impact Investing Network. $50,000
- JPMorgan Chase Foundation – Workforce Data Initiative $300,000 2016-2018
- Alfred P Sloan Foundation – Workforce Data Initiative $500,000 2016-2018
- NRT-IGE: Information Infrastructure for Society: Integrating Data Science and Social Science in Graduate Education and Workforce Development. Co-PI. $499,896
- Center for Aids Research, Pilot Grant, $50,000 2017
- Accenture – Support for Data Science for Social Good Conference, 2017. $100,000
- Alfred P Sloan Foundation. Data Science for Social Good Conference,2017. $35,000
- Accenture - Support for Workforce Development Research, 2018. $100,000
- City of Los Angeles - Data analytics for Anti-Sex and Labor Trafficking Unit (ASALT) of the Los Angeles City Attorney, 2019
- NSF - Community-Driven Data Engineering for Substance Abuse Prevention in the Rural Midwest, 2018-2020
- Joyce Foundation – National Police Intervention and Outcomes Research Consortium, 2018-2020
- C3.ai DTI award for artificial intelligence (AI) techniques to mitigate the COVID-19 pandemic. 2020 with Aziz Huq, Kit Rodolfa, Ryan Tibshirani $200,000
- Block Center for Technology and Society Grant with Kasun Amarasinghe, Ameet Talwalkar, Kit Rodolfa - $60k
- Beecher Family Foundation Gift $100,000
- NSF FAI: Fair AI in Public Policy - Achieving Fair Societal Outcomes in ML Applications to Education, Criminal Justice, and Health and Human Services with Hoda Heidari, Alex Chouldechova, and Zack Lipton - $600,000
- RK Mellon Foundation - Empowering Access for Social Equity with Shawn Blanton - $250,000
- Google AI for Social Good Grant, 2021-2022.

Boards
- Board of Directors, ChangeLab Solutions (Non-profit focused on Public Health Law and Policy)
- Board of Directors, Coleridge Initiative
- Board of Directors, Data Science for Social Good Foundation
- Board of Directors, AI for Good Foundation
- Advisory Committee, Data Science for Social Impact Washington University
- Steering Committee, Collaboratory Against Hate, CMU and University of Pittsburgh
- International Expert Panel, Australian Research Council Center for Information Resilience
- Advisory Council, Metrolab Network
- Advisory Board, Value based Healthcare Initiative at the World Economic Forum
- Board of Regents, Sewanee: The University of the South. 2019-2024
- Board of Trustees, Sewanee: The University of the South. 2017-2019
- Steering Committee, AI for Good Foundation
- NASPAA (Network of Schools of Public Policy, Affairs, and Administration) Research Committee
- Initiative for Analytics and Data Science Standards (IADSS) Advisory Board
- Technical Advisory Board, Hispanic Scholarship Fund (Non-profit)

- Advisory Board, Elastica (Network Security Startup)
- Alumni Board, Sewanee: The University of the South. 2010-2012
- Advisory Board, European Union funded Project on Semantically Enabled Knowledge Technologies (SEKT) 2004-2006

Invited Panels

- Atlantic Council Panel on "Importance of Human-Center Values and Fairness in AI Technologies" (May 2022)
- GDDF Panel on AI and Fairness. May 2022.
- "Building the NIST AI Risk Management Framework" Panel – organized by NIST. March 2022.
- "Can we Trust AI" Panel at the Institute for Experiential AI at Northeastern University (April 2022)
- Thoughtspot Data for Good 2019
- Voter protection Panel 2020
- DSAA 2019
- JPMC Institute Panel
- Urban Institute Panel
- Metrolab Conference, Oct 2018
- Innovation for Inclusion Summit: Harnessing Technology to Create More Equitable Cities, June 2018
- Artificial Intelligence and Accountability. BSA Annual Meeting, CA. November 2017
- Machine Learning: Gold Standards, Replicability, Reproducibility. Midwest Big Data Hub All Hands Meeting. October 2017
- 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
- Data Science for the Social Good. 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

Keynote Talks at Conferences

- Tredence 2019
- MIDAS
- IAPA Australia Oct 2018
- ACM SIGIR Conference, 2018
- Behavioral Exchange 2017, Singapore
- 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

- Invited talk at Applied AI & Diplomacy Series - organized by the US Department of State. September 2022
- Prato Summit on AI for Social Good – organized by Monash University Data Futures Institute. Prato, Italy. September 2022.
- USC Center for AI in Society – April 2022
- GAO Forum on AI oversight
- Tredence
- Invited talk at the American Association for Public Opinion Research (AAPOR) webinar. Aoril 2020.

- Value of science workshop
- Case Conference 2020
- Illinois Ignite September 2018
- Southside Civic Scopeathon, October 2018
- Invited talk on AI and Ethics October - Naspers. 2018
- Amazon Research Days (Machine Learning for Social Good) Nov 2018
- InterAmerican Development Bank - Machine Learning for Development Event, Nov 2018
- Data Scince Summer School , Porto, July 2018
- DataPub Meetup. Mexico City, 2017.
- ITAM, Mexico City, 2017
- Universidad Adolfo Ibanez, Santiago, Chile
- Challenge Week Keynote Talk, University of Central Arkansas. October 2017
- 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, Ocrtober 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 PKDDD 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/PKDDD 2003)
- Google Research
- Yahoo Research
- Toyota Technological Institute - Chicago
- DePaul University
- Notre Dame University

Events Organized:
- 2nd Data Science for Social Good Conference. 2017
- 1st Data Science for Social Good Conference. 2016
- 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

- Co-Director, Responsible AI Initiative at Carnegie Mellon University. 2022-present
- Working Group for "Ethical Use of Novel HIV Data and Analytics" at Georgetown University, funded by the Gates Foundation. September-November 2022
- Program Chair – Applications. IEEE Conference on Data Science and Advanced Analytics. 2018
- Host Committee, Computation + Journalism Symposium, 2017
- Social Impact Chair, KDD 2016, KDD 2017, KDD 2018, KDD 2020
- ACM SIGKDD Dissertation Award Committee, 2017, 2021
- ACM KDD Best Paper Committee, 2017
- Senior Program Committee Member, Applied Data Science Track, KDD 2016, 2017, 2018
- 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)
- International Conference on Machine Learning (ICML).