

Haimonti Dutta, Ph.D.

Assistant Professor

Department of Management Science and Systems
University at Buffalo, Buffalo, NY 14260
Office: 716-645-3259, Cell: 484-432-1484
Email: haimonti@buffalo.edu
Web: <http://www.buffalo.edu/~haimonti/>

Research Interests

Data Mining and Machine Learning, Distributed Optimization, Distributed and Parallel Data Mining, Probabilistic Inference, Machine Learning in Crowdsourcing, Ubiquitous and Data Intensive Computing, Big Data, Computational Neuroscience.

Applications: Healthcare, Digital Humanities, and Smart Environments (smart electrical grid, vehicles, buildings, wearable sensors).

Education

Jadavpur University, India	Computer Science and Engineering.	B.C.S.E. 1999
Temple University, Philadelphia, USA	Computer and Information Science.	M.S. 2002
University of Maryland, Baltimore County, USA	Computer Science and Electrical Engineering.	Ph.D. 2007

Dissertation: “Empowering Scientific Discovery by Distributed Data Mining on the Grid Infrastructure”
Advisor: Dr. Hillol Kargupta

Professional Experience

Aug.,2014-Present	Department of Management Science and Systems University at Buffalo, Buffalo, New York. Assistant Professor
Jun. 2016-Present	Computational and Data-Enabled Science and Engg. University at Buffalo, Buffalo, New York. Core Faculty Member
Jan.,2013-Jun. 2016	Department of Computer Science and Engineering Indraprastha Institute of Information Technology (IIIT), Delhi. Adjunct Assistant Professor.
Jan.,2012-Jun. 2016	Health Analytics and Foundations of Data Science Center(s) Data Science Institute Columbia University, New York, NY. Affiliated Member.
Sept.,2007-Mar.,2014	The Center for Computational Learning Systems, Columbia University, New York. Associate Research Scientist.
Jan.-May, 2011, 2012	Department of Computer Science, Columbia University, New York. Adjunct Assistant Professor.
Jan.,2003-Jul.,2007	University of Maryland, Baltimore County, MD. Dissertation Fellow, Graduate and Teaching Assistants over several semesters.
May-Aug.,2004	International Business Machines (IBM), T. J. Watson Research Laboratories, Media Delivery Architectures Department, Hawthorne, NY. Research Intern.
Aug.,2000-Jul.,2002	Temple University, Philadelphia. Graduate Research Assistant.
Aug.,1999-Jul.,2000	iGate Global Solutions, Chennai, India. Software Consultant.

Past and Current Research Support

- “Detection of Prejudice from Social Media Streams”
Funding Agency: School of Management, UB. Summer Research Award.
Amount: \$13,300
Period: Aug. 2018 - Apr. 2019.
PI: **Dr. Haimonti Dutta.**
- “Acquisition of High Performance Computing Infrastructure to Support Computational and Data-Enabled Science and Engineering”
Funding Agency: National Science Foundation, MRI
Period: Aug. 2017- Aug. 2018.
Amount: \$1,428,428
Role: Senior Personnel.
- “Consensus-based Machine Learning”
Funding Agency: School of Management, UB. Summer Research Award.
Amount: \$11,500
Period: Aug. 2017 - Apr. 2018.
PI: **Dr. Haimonti Dutta.**
- “Hadoop on Lake Effect Cloud, UB”
Funding Agency: Center for Computational Research (CCR), UB
Amount: Free subscription to Lake Effect Cloud for 120 students
Period: Aug. 2016 - Dec. 2016.
PI: **Dr. Haimonti Dutta.**
- “North East Big Data Innovation Hub: Health Spoke”
Funding Agency: National Science Foundation.
Period: Nov 2015-Oct 2016.
Role: Senior Personnel.
- “EEGMine: A Distributed Framework for Learning on EEG Data obtained from Epilepsy Patients”
Funding Agency: National Science Foundation, IIS-0916186
Amount: \$ 440,000 for 2 years, extended with no cost extension
Period: Aug 2009 - Aug 2013.
PI: **Dr. Haimonti Dutta**, Co-PIs: Dr. David Waltz, Dr. Catherine A. Schevon and Dr. Ronald Emerson.
- “Using Machine Learning to Understand the Scaling Behavior of the GFDL FMS High-Resolution Atmosphere Model on the Argonne BG/Q Platform”
Funding Agency: Argonne National Labs.
Amount: \$ 43,000
Period: 1 year.
PI: Dr. Jeffrey Durachta, Co-PIs: **Dr. Haimonti Dutta**, Robert Millstein
- “Development of distributed algorithms for incremental sensing and communication”
Funding Agency: EMC², India.
Amount: \$20,000 \approx INR 8 lakhs
Period: July 2013 - June 2014.
PI: Dr. Amarjeet Singh, Co-PI: **Dr. Haimonti Dutta**
- “An *early warning* device to allow epilepsy patients to lead a more normal life”
Funding Agency: Research Initiatives in Science and Engineering (RISE), Columbia University, New York.
Amount: \$ 170,000.
Period: Apr 2008 – Apr 2010.
PI: Dr David Waltz, Co-PIs: Dr. Catherine Schevon, Dr. Ron Emerson, Dr. Gail Kaiser, **Dr. Haimonti Dutta**, Dr. Ansaf Salleb-Aouissi
- “Intracranial EEG Acquisition System with Online Fast Ripple Detection”
Funding Agency: Epilepsy Research Foundation
Amount: \$200,000

Period: Sept 2009 – June 2012.

PI: Dr Catherine A Schevon, Co-PIs: Dr David Waltz, **Dr Haimonti Dutta**

- “A Distributed Framework for Learning on EEG data Obtained from Epilepsy Patients”
Funding Agency: Amazon Web Services (AWS) in Education Research Grant, Amazon, Inc.
Amount: \$3000 in infrastructure for use of the Amazon Elastic Compute Cloud (EC2), Amazon Simple Storage Service (S3), AWS Data Transfer and Amazon Virtual Private Cloud
Period: Jan 2010 - Dec 2010.
PI: **Dr. Haimonti Dutta**
- “Leveraging “The Wisdom of the Crowds” for Efficient Tagging and Retrieval of documents from the Historic Newspaper Archive of the New York Public Library”
Funding Agency: National Endowment of Humanities, Digital Humanities Start-Up Grant
Amount: \$49,452
Period: Sept, 2010 – Feb, 2013.
PI: **Dr. Haimonti Dutta**, Co-PIs: Dr. Rebecca Passonneau and Dr. David Waltz
- “Adaptive Stochastic Controller Load and Source Optimization For the Secure Interoperable Smart Grid Demonstration”
Funding Agency: Department of Energy
Amount: \$3 million
Period: June,2011 - Feb, 2012.
PI: Dr. Roger Anderson, Co-PIs: Dr. Doug Riecken, **Dr. Haimonti Dutta**, Dr. Ansaif Salleb-Aouissi
- “Conversion from Susceptibility Ranking to Mean-Time- Between-Failure Statistical Estimation for Feeders, Sections, and Joints”
Funding Agency: Consolidated Edison Company of New York
Amount: \$461,322
Period: July, 2008 - July 2009
PI: Dr. Roger Anderson, Co-PIs: Dr. David Waltz, Dr. Alessandro Moschitti, **Dr. Haimonti Dutta**.
- “Manhole Events and Secondary System Machine Learning Project - Phase 3 and 4”
Funding Agency: Consolidated Edison Company of New York
Amount: \$440,00 and \$413,947
Period: Jan - Dec, 2008.
PI: Dr. Roger Anderson, Co-PIs: Dr. Cynthia Rudin, **Dr. Haimonti Dutta**, Dr. Rebecca Passonneau.
- “Distributed and Peer-to-Peer Data Mining for Scalable Analysis of Data from Virtual Observatories”
Funding Agency: NASA, Research Opportunities in Earth and Space Science (ROSES)
Amount: \$722,541
Period: 2007–2011
PI: Dr. Hillol Kargupta, Role: Senior Personnel.
- Workshop Proposal: “Indo-Australia Workshop on Modeling Large-Scale Linked Data”
Funding Agency: Department of Science and Technology, India
Amount: INR 468,000
Period: November, 2013.
PI and Co-PIs: Dr. Ashwin Srinivasan, Dr. Mayank Vatsa, Dr. Richa Singh, Dr. Maya Ramanath, Dr. Parag Singla, Dr. Manik Varma, Dr. Gautam Shroff, **Dr. Haimonti Dutta**, Dr. Indrajit Bhattacharya, Dr. Srikanta Bedathur, Dr. Ganesh Ramakrishnan.

Honors and Awards

- Decision Support Systems Paper, “Intergroup Prejudice Detection: The Case of Microblogging under Terrorist Attacks”, Nominated for **Best Cybersecurity Paper** in National Security Agency’s, Science of Security (SoS) Paper Competition, 2019 [Currently Under Review]

- Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, “NIMLTK v0.2: A non-intrusive load monitoring toolkit for large scale data sets”, **Winner Best Demo Paper Award**, 2014.
- ICMLA paper on “Distributed Optimization Strategies for Mining on Peer-to-Peer Networks” nominated for the **Best Paper Award**, 2008.
- Dissertation Fellowship Award, UMBC Graduate School, Spring 2007.
- IBM Travel Grant awarded for participating at SIAM International Conference on Data Mining, Minneapolis, MN, April 2007.
- Graduate Research Association (UMBC) Travel Grant, April 2007.
- Travel Grant awarded by I.B.M Research for participating in Ph.D. Student Symposium, International Conference on Service Oriented Computing (ICSOC), Chicago, December 2006.
- Travel Grant awarded jointly by National Science Foundation (NSF) and Computing Research Association for Women (CRA-W) to participate in Workshop for Women in Machine Learning, San Diego, CA, 2006.
- Scholarship from BioInformatics Research Center (BRC), University of Maryland Baltimore County, Spring 2003.
- Ranked within top 125 (amongst 10,000 students) in West Bengal Joint Entrance Examination (WB JEE), 1995.
- Dr. B.C. Roy Scholarship for outstanding academic achievement, 1993 - 1994.

Presentations, Workshops and Invited Talks

- “Person Name Disambiguation Based on Profession”, 13th Data Mining and Decision Analytics Workshop, INFORMS Annual Meeting, Phoenix, AZ, Nov. 2018 (Also presented at “Statistical Data Analytics” session of the annual meeting.)
- “Optimization Methods for Machine Learning”, a three-day lecture series for Computer Science students, BITS Pilani, Goa, India. Aug. 2018.
- “Ranking Clinical Trials using Elasticsearch”, Text REtrieval Conference (TREC), Precision Medicine Track, Gaithersburg, MD, Nov. 2017.
- “Optimization Methods for Machine Learning”, 2-hr. Workshop at Computational and Data-Enabled Science and Engineering (CDSE) Days, UB, Apr. 2017.
- “PARABLE: A Parallel Random-Partition Based Hierarchical Clustering Algorithm for the MapReduce Framework”, INFORMS Annual Meeting, Nashville, TN, Nov. 2016.
- “Consensus-based Modeling Using Distributed Feature Construction by ILP”, INFORMS Annual Meeting, Nashville, TN, Nov. 2016.
- “Evaluation of Spell Correction on Noisy OCR Data.”, INFORMS Workshop on Data Mining and Analytics at INFORMS Annual Meeting, Philadelphia, October 2015.
- “Tackling the data deluge in epilepsy”, Management Science and Systems Department, School of Management, State University of New York at Buffalo, September 2013.
- “Tackling the data deluge in epilepsy”, Siemens Corporate Research, Princeton, NJ, August, 2013.
- “Large-scale Support Vector Machines: Current Research Trends and Future Directions”, Big Data School, Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Sydney, Australia, April 10th, 2013.
- “GADGET SVM: a Gossip-based sub-Gradient SVM Solver”, Tata Consultancy Services (TCS) Research Lab, New Delhi, April 2013.
- “Leveraging Collective Annotation for Machine Learning on the NYPL Historic Newspaper Archive”, Newspaper Interest Group, Annual Conference of the American Library Association, Anaheim, CA 2012.
- “Large-scale machine learning on intra-cranial electroencephalogram (iEEG) from epilepsy patients”, Department of Information Systems, Oakland University, Detroit, April 2012.
- “Leveraging Collective Annotation for Machine Learning on the NYPL Historic Newspaper Archive”, co-hosted by the Department of Library Science and Computer Science and Engineering, University of North Texas, Denton TX, April 2012.
- “Supervised Learning on Peer-to-Peer Networks”, Invited Talk, Accenture Technology Labs, Reston VA, 2012.
- “Large-scale machine learning on intra-cranial electroencephalogram (iEEG) from epilepsy patients”, Department of Information Systems, NJIT, March 2012.
- “Leveraging Collective Annotation for Machine Learning on the NYPL Historic Newspaper Archive”, IIT Delhi, March, 2012.
- “A Case-Study on Learning from Large-scale Intracranial EEG Data using Multi-core Machines and Clusters”, The Third Workshop on Large-scale Data Mining: Theory and Applications, SIGKDD, San Diego, August, 2011.
- “Learning Parameters of the K-Means Algorithm from Subjective Human Annotation.”, The 24th International FLAIRS Conference, Palm Beach, FL. May 18-20, 2011.
- “Modeling Failures of Electrical Components and Power Flow in Smart Power Grids: Progress and Challenges”, Santa Fe Institute, NM, Workshop on Decentralized Control in Systems of Strategic Actors, Invited Talk, August 2010.

- “Algorithms for Distributed Supervised and Unsupervised Learning”, NY Machine Learning Group, AOL Headquarters, NY, April, 2010.
- “Data Mining on Intracranial EEG obtained from Epilepsy Patients”, Trenton Computer Festival (TCF) at the College of New Jersey, 2010.
- “Measuring Diversity of Regression Ensembles”, 4th Indian International Conference on Artificial Intelligence (IICAI), Tumkur, India, December, 2009.
- “Distributed Data Mining on Intracranial EEG obtained from Epilepsy Patients”, IBM Research, Bangalore, India, Invited Talk, December, 2009.
- “Machine Learning: An Introduction”, Trenton Computer Festival (TCF) at the College of New Jersey, 2009.
- “Distributed Linear Programming and Resource Management for Data Mining in Distributed Environments”, 10th International Workshop on High Performance Data Mining (HPDM) held in conjunction with the International Conference on Data Mining (ICDM), Pisa Italy.
- “Machine Learning Susceptibility Ranking System”, Special International Utility Working Group(IUWG) Workshop on Computer-Aided Lean Management for the Energy Industry, April, 2008.
- “Distributed Optimization Strategies for Mining on Peer-to-Peer Networks”, ICMLA, San Diego, December 2008.
- “Visualization of Manhole and Precursor-Type Events for the Manhattan Electrical Distribution System”, Workshop on Geo-Visualization of Dynamics, Movement and Change, Girona, Spain, May 2008.
- “Distributed Top-K Outlier Detection from Astronomy Catalogs using the DEMAC System”, SIAM International Conference on Data Mining, Minneapolis, MN USA, April 2007.
- Talk on my Ph.D. thesis, I.B.M. Ph.D. Student Symposium at International Conference on Service Oriented Computing (ICSOC), Chicago, December, 2006.
- “Distributed Data Mining on Astronomy Catalogs”, Poster Presentation at the Workshop for Women in Machine Learning (WiML), San Diego, October 2006.
- “Empowering Scientific Discovery by Distributed Data Mining on the Grid Infrastructure”, Ph.D. Proposal Defense Presentation, Department of Computer Science and Electrical Engineering, UMBC, August 2006.
- “A Routing Algorithm for Content Based Publish Subscribe Data Streams”, I.B.M, T.J.Watson Research Laboratory, August 2004.
- “Classification of 3D Region Data”, The Center of Information Science and Technology, Temple University, October 2001.
- “Shape Representation and Matching in Medical Tumor Databases (using mathematical morphology)”, CIS Department, Temple University, April 2001.
- “Spatial Data Mining”, CIS Department, Temple University, April 2001.
- “A Tutorial on Bayesian Networks”, Temple University, Nov 2000.
- “Maintenance of Software”, Jadavpur University, Kolkata, India, April 1998.

Software

- **Gossip-bAseD sub-GradiEnT SVM Solver (GADGET SVM)**
WebSite: <https://github.com/nitinnat/GADGET>
 - A consensus based learning algorithm for distributed environments (such as wireless sensor networks, mobile applications and ubiquitous computing environments)
 - Version 1.0 of the software built using the Peersim simulator (<http://peersim.sourceforge.net/>)

- Targeted applications: mobile applications for libraries, non-intrusive load monitoring for efficient energy utilization in buildings.
- Forthcoming, extension of the algorithm to dynamic networks, studying related gossip protocols, examination of the relation between the graph structure and its impact on convergence of the algorithm).

- **Density Estimation based Ranking from Decision Trees**

WebSite: <http://www1.ccls.columbia.edu/~dutta/Software/PET.html>

- **Fourier Representation of Decision Trees**

WebSite: <http://www.csee.umbc.edu/~hdutta1/FSWebsite/index.html>

Decision tree ensembles are frequently used in data mining and machine learning applications. The Fourier representation of decision trees has several interesting properties that are particularly useful for aggregating, transmitting and visualizing decision trees. The main motivation behind this approach is to create an algebraic framework for meta-level analysis of models, produced by many ensemble learning, data stream mining, distributed data mining, and other related techniques. The software developed allows the data miner to convert a decision tree to its Fourier Spectrum and vice-versa. It is open-source and has appropriate documentation illustrating how it can be used. The data mining community has expressed keen interest in its usage and the software is also available for download from the free encyclopedia for Distributed and Ubiquitous Data Mining (DDMWiki) http://www.umbc.edu/ddm/wiki/index.php/Main_Page.

Academic Teaching Experience

- MGS 647 TUT - S1S, Supervised Research, Spring 2018 [3 students], Fall 2017 [2 students], Spring 2017 [1 student]. Students from Computer Science, Data Science and Management Information Systems were enrolled.
- M.S. in Management Information Systems (MIS) Practicum Supervision, Spring 2017, 2018.
- MGS 798, Tutorial, MSS Workshop for 1st - 2nd year Ph.D. students, Fall 2018 [Class size: 4, Seminar participants: 20+]
- “Convex Optimization for Machine Learning”, MGS 662 (3 credits), Assistant Professor, Department of Management Science and Systems, University at Buffalo, Spring 2019 [Class Size: 10]
- “Machine Learning for IT Managers”, MGS 662 (3 credits), Assistant Professor, Department of Management Science and Systems, University at Buffalo, Spring 2017 [Class Size: 12 students], 2018 [Class Size: 29 students].
- “Machine Learning for IT Managers”, MGS 618 (1.5 credits), Assistant Professor, Department of Management Science and Systems, University at Buffalo, Spring 2015, 2016.
- “Big Data Information Management”, MGS 618 (1.5 credits), Assistant Professor, Department of Management Science and Systems, University at Buffalo, Spring 2015, 2016.
- “Distributed Computing and Big Data Technologies”, MGS 655 (3 credits), Assistant Professor, Department of Management Science and Systems, University at Buffalo, Fall 2014, 2015, 2016, 2017, 2018 [Class Size: 120+ students].
- “Distributed Data Mining”, CSE 608, Assistant Professor, Department of Computer Science, IIIT-Delhi, India, Spring 2013 [Class Size: 45].
The following is a summary of student evaluations for the above course (no of responses=9, total number of students in the class = 15) [1: Not Helpful, 5: Excellent]
- “An Introduction to Distributed Data Mining”, COMS 6998, Section 009, Adjunct Faculty, Department of Computer Science, Columbia University, New York, Spring 2012.
- “An Introduction to Distributed Data Mining”, COMS 6998, Section 006, Adjunct Faculty, Department of Computer Science, Columbia University, New York, Spring 2011.

	Question	Mean
1.	Course: Achieved learning outcomes	4
2.	Course: Homework/Assignment appropriateness	4.78
3.	Course: Usefulness of project	4.56
4.	Course: Use of textbooks and references	3.44
5.	Course: Grading Policy	4.56
6.	Instructor: Facilitated learning	3.67
7.	Instructor: Lectures were helpful	3.22
8.	Instructor: Effective in-class interactions	3.44
9.	Instructor: Tutorials / Office-hour interaction	4.33
10.	Overall: Score	4.09

- Teaching Assistant for CMSC 441 (Algorithm Design and Analysis) and CMSC 461 (Database Management Systems), University of Maryland, Baltimore County.
September 2006 - December 2006
- Supervisor, Math and Science Resource Center (MSRC), Temple University. Supervised and tutored undergraduate students in Mathematics and Statistics
July 2001 - September 2002
- Tutor, Kolkata, India. Tutored High School Students in Mathematics, Statistics, Physics, Chemistry and Biology.
June 1997 - June 1999

Professional Service

Program Committees

- Neural Information Processing Systems (NIPS 2012, 2013, 2014, 2015, 2016, 2017).
- International Conference on Machine Learning (ICML 2012, 2013, 2014, 2017, 2018, 2019).
- AI & Statistics (AISTATS 2014).
- IEEE Big Data 2013.
- International Conference on Learning Representations, 2017.
- International Conference on Advanced Computing, Networking and Informatics, 2018.
- Second International Conference on Big Data Analytics (BDA-2013).
- CIKM Workshop on Data Management and Analytics for Healthcare (DARE, 2013).
- Euro-Par, Topic PC Member, (2005, 2012).
- International Conference on Machine Learning and Applications (ICMLA 2010, 2011).
- International Conference on Advances in Social Networks Analysis and Mining (ASONAM, 2010, 2011, 2012, 2013).
- International Conference on Data Mining (ICDM 2006, 2009, 2010, 2011, 2012, 2013).
- SIAM International Conference on Data Mining (SDM 2005, 2006, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016).
- ACM Conference on Knowledge Discovery and Data Mining, (ACM SIGKDD 2007, 2008, 2009, 2010, 2011).
- Pacific-Asia Knowledge Discovery and Data Mining, (PAKDD 2008, 2009, 2011, 2012, 2013).
- ACM Fifteenth Conference on Information and Knowledge Management, (CIKM 2006).
- Indian International Conference on Artificial Intelligence, (IICAI 2011).

- European Conference on Machine Learning and Practice of Knowledge Discovery in Databases (ECML-PKDD 2008).
- Annual Workshop for Women in Machine Learning (WiML 2007)
- Grace Hopper Celebration of Women and Computing, Travel Scholarships (2008, 2009, 2010).

Reviewing Proposals

- Panelist for NSF III Program, 2013; NSF CDI Program, 2010; NSF IIS Program, 2008.
- Panelist for Portuguese Foundation for Science and Technology (FCT), 2012.
- Panelist for The Council for Physical Sciences of the Netherlands Organization for Scientific Research (NWO), 2005.

Journals

- Decision Support Systems, 2018.
- IEEE Tran. on Computers, 2018.
- IEEE Tran. on Smart Grid, 2018.
- Coordinating Editor, Information Systems Frontiers (2015).
- INFORMS, Journal of Computing (2014, 2015).
- IEEE Transactions on Big Data (2015, 2016).
- IEEE Transactions on Systems, Man and Cybernetics (2010, 2008).
- Data Mining and Knowledge Discovery (DMKD) Journal (2009, 2012, 2013).
- WIRES Data Mining and Knowledge Discovery 2009.
- IEEE Transactions on Knowledge and Data Engineering (TKDE) (2007, 2008, 2011, 2016).
- Adhoc Reviewer, Special Issue of Information Fusion Journal

Research Advising / Supervision

• Doctoral

- Srikanth Parameswaran, Ph.D., UB (Role: Dissertation Committee Member), Graduated, 2018.
- Suchismit Mahapatra, Ph.D., Department of Computer Science, UB (Role: Thesis Committee Member), Graduated, 2018.
- Chris Bang, Ph.D., UB. (Role: Dissertation Committee Member). Graduated, 2017.
- Shen Wang, Ph.D., Columbia University. (Role: Advisor). 2010-2013.

• Junior Colleagues

- Kevin Mc Inerny, Staff Associate, The Center for Computational Learning Systems, Columbia University. (2011-2012).
- Ashish Tomar, Staff Associate, The Center for Computational Learning Systems, Columbia University. (2012).
- Axinia Radeva, Staff Associate, The Center for Computational Learning Systems, Columbia University. (2007-2014).
- Boyi Xie, Ph.D. student, Columbia University.(2011-2012).
- Somnath Sarkar, Department of Industrial Engineering and Operations Research, Columbia University. (2012).

• Master's Theses Supervised

- Aayushee Gupta, “Finding Influential People from a Historical News Archive”, M.Tech, Data Engineering, IIIT Delhi (Role: Advisor, Thesis Committee Members: Dr Srikanta Bedathur, Lipika Dey).
- Swati Agrawal, “Detecting Copyright Violation on YouTube Videos using YouTube Meta-Data”. M.Tech, Data Engineering, IIIT Delhi (Role: Thesis Committee Member).
- Megha Gupta, “Evaluation of OCR Text Correction By Crowdsourcing on a Historic Newspaper Archive”, M. Tech, Data Engineering, IIIT Delhi (Role: Advisor).

• Master’s Students

- Neha Lalaso Jagtap, M. S. in Computer Science, UB. 2018 (Role: Supervisor, Independent Study).
- Jayashree Chandrasekaran, M. S. in Computer Science, UB. 2018 (Role: Supervisor, Independent Study).
- Nitin Nataraj, M. S. in Computer Science, UB. 2018 (Role: Supervisor, Independent Study).
- Kaushik Panneerselvam, M. S. in Computer Science, UB. 2018 (Role: Supervisor, Independent Study).
- Sushma Sharma, M.S. in MIS, UB, 2018. (Role: Supervisor of Independent Study).
- Akshat Shehgal, M.S. in MIS, UB, 2017. (Role: Supervisor of M.S. Practicum).
- Samridhi Sharma, M.S. in MIS, UB, 2017. (Role: Supervisor of M.S. Practicum).
- Mihir Chauhan, M.S. in Computer Science, UB. 2017 (Role: Supervisor, Independent Study).
- Aayushee Gupta, Data Engineering, IIIT-Delhi, India (Role: Advisor). 2013-present.
- Lovey Agrawal, Data Engineering, IIIT-Delhi, India (Role: Advisor). 2013-2014.
- Deepak Nayak(2012-2013), Suraj Kesari (2012), William Chan(2010-present), Deepak Shakargouda(2011-2012), Nandini Bhardwaj(2007-2008), Ananda Mathur(2007-2008), Haoyun Feng(2008), Devendra Laukar(2008), Cheng Cheng(2008), Karthik Marudachallam Ramasamy(2008-2011), Huascar Fiorletta(2009-2010), Vinodkumar Prabhakaran(2008-2009, Co-supervised with Dr. Rebecca Passonneau), Abhinav Saini(2008), Mariya Riskova(2008), Austin Lee(2010-2012), Taylor Brown(2009), Yash Satsangi(2011-2012) – Department of Computer Science, Columbia University, NY.
- Connie Lee(2010), Pascal Notin(2011), John Zhang(2011-2012), Raghuram Nagireddy(2012-present) – Industrial Engineering and Operations Research, Columbia University.
- Tushar Mahule(2011-2012), Xianshu Zhu(2011-2012), – University of Maryland, Baltimore County, MD (Co-advisor Dr. Hillol Kargupta).
- Amit Sengupta(2013) – Department of Biostatistics, Columbia University.
- Vashist Avadhanula(2011-present), Dual Degree (B.Tech and M.Tech), Department of Electrical Engineering, IIT-Bombay.
- Priyanshu Jain(2012-present), Dual Degree (B. Tech and M. Tech), Department of Electrical Engineering, IIT-Delhi.

• Undergraduate Students

- Chase Hensel(2008-2010), Stanley German(2010), Margret Una Kjartansdottir, Stephen Pratt, Kyle Rego – Department of Computer Science, Columbia University, NY.
- Alex Kamil(2010-2011), School of General Studies, Columbia University.

• High School Students

- Shoshana Gottesman, Touro College, Funded by NSF DREU Program, 2010.
- Rivka Levitan, Brooklyn College, CUNY, Funded by NSF DREU Program, 2010.
- Ashutosh Nanda, Senior Montgomery County High School, 2013-present.

Teaching Service

- Ph.D. Program Faculty Advisor, MSS Department, UB 2018-2019.

- Member of Ethics Committee, MSS Department, UB, 2018.
- Member of Master's thesis award's committee, Department of Computer Science, IIIT-Delhi.

Colloquium

- Machine Learning Brown Bag Lunch, The Center for Computational Learning Systems, Columbia University. (2011-2012, 2007-2008).

Affiliations

Member of:

- Center for Data Science and Engineering (CDSE), University at Buffalo, Buffalo, New York.
- Special Interest Group in Relational Learning (SIGRL), IIIT-Delhi, India.
- DataLab, IIIT-Delhi, India (Research Group comprising of faculty members from IIT, IIIT and TCS Innovation Research Labs, Delhi).
- Association for Computing Machinery (ACM).
- INFORMS.
- Institute of Electrical and Electronics Engineers (IEEE).
- Center for Women in Technology, UMBC; Center for Women In Technology (CWIT), Columbia University, NY.

JOURNAL ARTICLES

- **Haimonti Dutta** and Ashwin Srinivasan. “Consensus Based Modeling Using Distributed Feature Construction with ILP”, *Machine Learning*, 107 (5), 825-858, 2018.
URL: <https://link.springer.com/article/10.1007/s10994-017-5672-2>
- **Haimonti Dutta**, Hazel Kwon, and H. Raghav Rao. “Intergroup Prejudice Detection: The Case of Microblogging under Terrorist Attacks”, *Decision Support Systems*, Vol. 113, pp 11-21, 2018.
URL: <https://www.sciencedirect.com/science/article/pii/S0167923618301003>
Media Impressions:
Campus Reform: <https://www.campusreform.org/?ID=11423>,
UBNow: <http://www.buffalo.edu/ubnow/research/news.host.html/content/shared/university/news/ub-reporter-articles/stories/2018/10/dutta-prejudice-social-media.detail.html>
- **Haimonti Dutta**. “A Consensus Algorithm for Linear Support Vector Machines”, Revise and Resubmit, *Management Science*, Jul. 2019.
URL: <https://arxiv.org/abs/1812.02261>
- **Haimonti Dutta** and Aayushee Gupta, “PNRank: A Reputation System for Person Name Entities Based on Unsupervised Ranking”, Submitted to *ACM Transactions on Management Information Systems*.
- Xianshu Zhu, Tushar Mahule, **Haimonti Dutta**, Sugandha Arora, Hillol Kargupta, Kirk Borne, “Peer-to-Peer Distributed Text Classifier Learning in PADMINI”, *Statistical Analysis and Data Mining Journal* 5 (5): 446-462, 2012.
- Cynthia Rudin, David Waltz, Roger N. Anderson, Albert Boulanger, Ansaif Salleb-Aouissi, Maggie Chow, **Haimonti Dutta**, Philip Gross, Bert Huang, Steve Ierome, Delna Isaac, Arthur Kressner, Rebecca J. Passonneau, Axinia Radeva and Leon Wu, “Machine Learning for the New York City Power Grid”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2011.
URL: <http://web.mit.edu/rudin/www/TPAMIPreprint.pdf>
- Cynthia Rudin, Becky Passanneau, Axinia Radeva, **Haimonti Dutta**, Steve Ierome and Delfina Isaac, “A Process for Predicting Manhole Events in Manhattan”, *Machine Learning*, Vol 80, Number 1, pages 1 -31, July, 2010.
URL: <http://www.springerlink.com/content/3034h0j334211484/>
- Hillol Kargupta, Byung Hoon Park, **Haimonti Dutta**, “Orthogonal Decision Trees”, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2006, volume 18, number 7, pp. 1028-1042.
URL: <http://www1.ccls.columbia.edu/~dutta/TKDE.pdf>

MANUSCRIPTS IN PREPARATION

- **Haimonti Dutta** and Ashwin Srinivasan, “A Consensus Algorithm for Distributed Models and Data with Application to Relational Data Streaming”, Targeted Journal: Machine Learning.
- **Haimonti Dutta**. “A consensus-based algorithm for learning distributed deep networks from vertically partitioned data”, Targeted Journal: IEEE PAMI.
- **Haimonti Dutta**. “Prejudice Detection from Streaming Data”, Targeted Journal: Management Science.
- Bao Hyunh, **Haimonti Dutta**, Dane Taylor, “On the convergence of a distributed support vector machine algorithm on modular networks”, Targeted Journal: JMLR.

¹H-index: 14, Citations: 884

BOOK CHAPTERS

- **Haimonti Dutta**, Alex Kamil, Manoj Pooleery, Simha Sethumadhavan and John Demme, “Distributed Storage of Large Scale Multidimensional Electroencephalogram Data using Hadoop and HBase”, In *Grid and Cloud Database Management*, Editors Sandro Fiore and Giovanni Aloisio, Springer, 2011
URL: <http://www1.ccls.columbia.edu/~dutta/gcdm.pdf>

PATENTS

- **Inventor: Haimonti Dutta.** Machine Learning for Power Grid, Patent Number: 8,751,421. Awarded: June, 2014.

REFEREED CONFERENCE AND WORKSHOP PAPERS

- **Haimonti Dutta**, “The effect of stochastic approximations on a Gossip-bAseD sub-GradiEnT solver for Linear SVMs”, 32nd International Florida Artificial Intelligence Research Society (FLAIRS) Conference, Sarasota, FL, 2019.
- **Haimonti Dutta**, “Games, Auctions and Consensus-based Machine Learning”, 32nd International Florida Artificial Intelligence Research Society (FLAIRS) Conference, Sarasota, FL, 2019.
- **Haimonti Dutta**, Jayashree Chandrasekaran, Kaushik Panneerselvam, “Person Name Disambiguation Based on Profession”, Workshop on Data Mining and Decision Analytics, INFORMS Annual Meeting, Nov. 2018.
- **Haimonti Dutta** and Ajinkya Thorve, “Ranking Clinical Trials Using Elasticsearch”, Proc. of the 26th Text REtrieval Conference (TREC), Precision Medicine Track, Gaithersburg, MD, Nov. 2017.
- Aayushee Gupta and **Haimonti Dutta**, “Evaluation of Spell Correction on Noisy OCR Data.”, INFORMS Workshop on Data Mining and Analytics at INFORMS Annual Meeting, Philadelphia, October 2015.
- Megha Gupta, **Haimonti Dutta**, Brian Geiger, “Classification of Crowdsourced Text Correction”, In Proc. of 2nd iKDD Conference on Data Sciences (CODS), Bangalore, India, Pages 142-143, Mar. 2015.
URL: <http://dl.acm.org/citation.cfm?doid=2732587.2732619>
- **Haimonti Dutta** and Ashwin Srinivasan, “Consensus-Based Modeling Using Distributed Feature Construction”, In Proc. of the 24th International Conference on Inductive Logic Programming (ILP), Sep. 2014, Nancy France.
- Jack Kelly, Nipun Batra, Oliver Parson, **Haimonti Dutta**, William Knottenbelt, Alex Rogers, Amarjeet Singh and Mani Srivastava, “NILMTK v0.2: A Non-Intrusive Load Monitoring Toolkit for Large Scale Datasets”, In Proc. of the 1st ACM Conference on Embedded Systems for Energy Efficient Buildings, Pages 182-183, 2014.
- Nipun Batra, Jack Kelly, Oliver Parson, **Haimonti Dutta**, William Knottenbelt, Alex Rogers, Amarjeet Singh and Mani Srivastava, “NILMTK: An Open Source Toolkit for Non-Intrusive Load Monitoring”, In Proceedings of the 5th International Conference on Future Energy Systems (ACM e-Energy), Cambridge, UK, Pages 265-276, Jun. 2014.
URL: <http://dl.acm.org/citation.cfm?id=2602051>
- Nipun Batra, **Haimonti Dutta** and Amarjeet Singh. “INDiC: Improved Non-Intrusive Load Monitoring using Load Division and Calibration”, Proc. International Conference on Machine Learning and Applications (ICMLA), Miami, FL, Pages 79-84, 2013.

- **Haimonti Dutta** and William Chan. “Using community structure detection to rank annotators when ground truth is subjective”, NIPS Workshop on Human Computation for Science and Computational Sustainability, Lake Tahoe, December 7th, 2012.
URL: <http://www1.ccls.columbia.edu/~dutta/nips2012.pdf>
- Rebecca J. Passonneau, Ashish Tomar, Somnath Sarkar, **Haimonti Dutta** and Axinia Radeva, “Multivariate Assessment of a Repair Program for a New York City Electrical Grid”, 11th International Conference on Machine Learning and Applications ICMLA, Special Session on Machine Learning in Energy Applications, Boca Raton, FL, Dec 13 - 15, 2012.
URL: http://www1.ccls.columbia.edu/~dutta/inspections_icmla_2012.pdf
- Boyi Xie, Rebecca J. Passonneau, **Haimonti Dutta**, Jing-Yew Miaw, Axinia Radeva, Ashish Tomar and Cynthia Rudin, “Progressive Clustering with Learned Seeds: An event categorization system for Power Grid”, In Proceedings of 24th International Conference on Software Engineering and Knowledge Engineering (SEKE 2012). Redwood City, CA. July 1-3, 2012.
URL: <http://www1.ccls.columbia.edu/~beck/pubs/seke2012.pdf>
- **Haimonti Dutta**, “A Randomized Gossip-based Algorithm for Classification on Peer-to-Peer Networks”, In Proceedings of the *NIPS Workshop on Big Learning: Algorithms, Systems, and Tools for Learning at Scale*, Grenada, Spain, Dec 2011.
URL: <http://www1.ccls.columbia.edu/~dutta/nips2011.pdf>
- **Haimonti Dutta**, Huascar Fiorletta, Manoj Pooleery, Hatim Diab, Stanley German, David Waltz, “A Case-Study on Learning from Large-scale Intracranial EEG Data using Multi-core Machines and Clusters”, In Proceedings of The Third Workshop on Large-scale Data Mining: Theory and Applications, SIGKDD, San Diego, August, 2011.
URL: <http://www1.ccls.columbia.edu/~dutta/sigproc-sp.pdf>
- **Haimonti Dutta**, Rebecca J. Passonneau, Austin Lee, Axinia Radeva, Boyi Xie, David Waltz and Barbara Taranto, “Learning Parameters of the K-Means Algorithm from Subjective Human Annotation.”, *The 24th International FLAIRS Conference, Special Track on Data Mining*, Palm Beach, FL. May 18-20, 2011.
URL: <http://www1.ccls.columbia.edu/~dutta/flairs.pdf>
- Leon Wu, Gail Kaiser, Cynthia Rudin, David Waltz, Roger Anderson, Albert Boulanger, Ansaf Salleb-Aouissi, **Haimonti Dutta**, and Manoj Pooleery, “Evaluating Machine Learning for Improving Power Grid Reliability”, Proceedings of the Workshop on Machine Learning for Global Challenges, International Conference on Machine Learning (ICML), 2011.
URL: <http://web.mit.edu/rudin/www/WuEtAlNOVA11.pdf>
- Shen Wang and **Haimonti Dutta**, “PARABLE: A PARallel RAndom-partition Based Hierarchical Clustering Algorithm for the MapReduce Framework”, 6th Annual Machine Learning Symposium at the New York Academy of Science (NYAS), 2011.
URL: http://www1.ccls.columbia.edu/~dutta/nyas2011_parable.pdf
- **Haimonti Dutta**, “Density Estimation Based Ranking from Decision Trees”, 6th Annual Machine Learning Symposium at the New York Academy of Science (NYAS), 2011.
URL: http://www1.ccls.columbia.edu/~dutta/nyas2011_PET.pdf
- **Haimonti Dutta**, David Waltz, Karthik M Ramasamy, Phil Gross, Ansaf Salleb-Aouissi, Hatim Diab, Manoj Pooleery, Catherine A Schevon and Ronald Emerson, “Patient-Specific Seizure Detection From Intra-cranial EEG Using High Dimensional Clustering”, *IEEE Proceedings of the International Conference on Machine Learning and Applications (ICMLA)*, Bethesda, MD, 2010.
URL: <http://www1.ccls.columbia.edu/~dutta/~icmla.pdf>

- Austin Lee, **Haimonti Dutta**, Rebecca J. Passonneau, David Waltz and Barbara Taranto, “Topic Identification from Historic Newspaper Articles of the New York Public Library: A Case Study”, 5th Annual Machine Learning Symposium at the New York Academy of Science, 2010.
- Margret Una Kjartansdottir, **Haimonti Dutta**, Catherine A Schevon, Ansaf Salieb-Aouissi, David Waltz and Ronald Emerson, “Detection of High Frequency Oscillations Using Support Vector Machines: A Case Study”, WiML, 2010 (Held in Conjunction with NIPS 2010), Vancouver, BC.
- **Haimonti Dutta**, David Waltz, Ansaf Salieb-Aouissi, Catherine Schevon and Ronald Emerson, “Designing Patient-Specific Seizure Detectors From Multiple Frequency Bands of Intra-cranial EEG Using Support Vector Machines”, Workshop on Data Mining for HealthCare Management (DMHM) held in conjunction with the 14th Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2010, Hyderabad, India.
URL: <http://www1.ccls.columbia.edu/~dutta/pakdd.pdf>
- Phil Gross, Ansaf Salieb-Aouissi, **Haimonti Dutta** and Albert Boulanger, “Ranking Electrical Feeders of the New York Power Grid”, IEEE Proceedings of the *International Conference on Machine Learning and Applications (ICMLA)*, pages 359 – 365, Miami, FL, December 2009.
URL: <http://www.computer.org/portal/web/csdl/doi/10.1109/ICMLA.2009.99>
- **Haimonti Dutta**, “Measuring Diversity in Regression Ensembles”, *Proceedings of the 4th Indian International Conference on Artificial Intelligence (IICAI)*, pages 2220- 2236, Bangalore, India, December 2009.
URL: <http://www1.ccls.columbia.edu/~dutta/iicai.pdf>
- Chase Hensel and **Haimonti Dutta**, “GERMS: a distributed sub-Gradient ERM Solver”, 4th Annual Machine Learning Symposium at the New York Academy of Sciences (NYAS), New York, November, 2009.
- **Haimonti Dutta**, David Waltz, Alessandro Moschitti, Daniele Pighin, Philip Gross, Claire Monteleoni, Ansaf Salieb-Aouissi, Albert Boulanger, Manoj Pooleery and Roger Anderson, “Estimating the Time Between Failures of Electrical Feeders in the New York Power Grid”, Next Generation Data Mining Summit (NGDM), Columbia, MD, October, 2009.
- **Haimonti Dutta**, Xianshu Zhu, Tushar Mahule, Hillol Kargupta, Kirk Borne, Codrina Lauth, Florian Holz, and Gerherd Heyer, “TagLearner: A P2P Classifier Learning System from Collaboratively Tagged Text Documents”, International Conference on Data Mining (ICDM), Workshop on Mining Multiple Information Sources, pages 495 – 500, Miami, December 2009.
URL: <http://www1.ccls.columbia.edu/~dutta/ICDMW.pdf>
- Chase Hensel and **Haimonti Dutta**, “GADGET SVM: a Gossip-bAseD sub-GradiEnT SVM Solver”, International Conference on Machine Learning (ICML), Numerical Mathematics in Machine Learning Workshop, Montreal, Quebec, 2009.
URL: <http://www1.ccls.columbia.edu/~dutta/GadgetSVM.pdf>
- **Haimonti Dutta**, David Waltz, Catherine A. Schevon, Karthik M Ramasamy, Phil Gross, Ansaf Salieb-Aouissi, Hatim Diab, Manoj Pooleery, Albert Boulanger and Ron Emerson, “Seizure Detection from Multiple Frequency Bands of Intra-cranial EEG using High Dimensional Clustering”, 4th International Workshop on Seizure Prediction, Kansas City, MO, June 4th - 7th, 2009.
URL: <http://www1.ccls.columbia.edu/~dutta/iwsp4.pdf>
- **Haimonti Dutta** and Ananda Matthur, “Distributed Optimization Strategies for Mining on Peer-to-Peer Networks”, IEEE Proceedings of the *International Conference on Machine Learning and Applications (ICMLA)*, pages 350 - 355, San Diego, 2008. **Nominated for the Best Paper Award**
URL: <http://www1.ccls.columbia.edu/~dutta/paperICMLA.pdf>

- Haoyun Feng, **Haimonti Dutta** and Ansaf Salleb-Aouissi, “On Improving Probability Estimate Trees”, 3rd Annual Workshop for Women in Machine Learning (WiML) held in conjunction with Neural Information Processing Systems (NIPS), Vancouver, B.C., 2008.
URL: <http://www1.ccls.columbia.edu/~dutta/WiML-abstract-5.pdf>
- Phil Gross, Ansaf Salleb-Aouissi, **Haimonti Dutta** and Albert Boulanger, “Ranking Electrical Feeders of the New York Power Grid”, 3rd Annual Machine Learning Symposium at the New York Academy of Sciences (NYAS), New York, October, 2008.
URL: <http://www1.ccls.columbia.edu/~dutta/extabstract.pdf>
- **Haimonti Dutta** and Hillol Kargupta, “Distributed Linear Programming and Resource Management for Data Mining in Distributed Environments”, 10th International Workshop on High Performance Data Mining (HPDM), Held in Conjunction with International Conference on Data Mining (ICDM), pages 543 – 552, Pisa, Italy, 2008.
- **Haimonti Dutta**, Cynthia Rudin, Becky Passonneau, Fred Seibel, Nandini Bhardwaj, Axinia Radeva, Zhi An Liu, Steve Ierome and Delfina Isaac, “Visualization of Manhole and Precursor-Type Events for the Manhattan Electrical Distribution System”, *Workshop on Geo-Visualization of Dynamics, Movement and Change*, 11th AGILE International Conference on Geographic Information Science, Girona, Spain, 2008.
URL: http://www1.ccls.columbia.edu/~dutta/Viz_08.pdf
- **Haimonti Dutta**, Chris Giannella, Kirk Borne, Hillol Kargupta, “Distributed Top-K Outlier Detection for Astronomy Catalogs using the DEMAC System”, *Proceedings of the Seventh SIAM International Conference on Data Mining*, Minneapolis, USA, 2007.
URL: <http://www1.ccls.columbia.edu/~dutta/SDM07.pdf>
- Haimonti Dutta, “Empowering Scientific Discovery by Distributed Data Mining on the Grid Infrastructure”, *IBM Symposium for PhD Students at 4th International Conference on Service-Oriented Computing*, 2006.
URL: <http://www1.ccls.columbia.edu/~dutta/ICSOC06.pdf>
- Chris Giannella, Haimonti Dutta, Sourav Mukherjee and Hillol Kargupta, “Distributed Kernel Density Estimation”, *The 9th International Workshop on High Performance and Distributed Data Mining (HPDM)*, 2006.
URL: <http://www1.ccls.columbia.edu/~dutta/HPDM06.pdf>
- Chris Giannella, Haimonti Dutta, Ran Wolff, Kirk Borne and Hillol Kargupta, “Distributed Data Mining in Astronomy Databases”, *The 9th Workshop on Mining Scientific and Engineering Data Sets (held in conjunction with Siam Data Mining Conference)*, 2006.
URL: http://www1.ccls.columbia.edu/~dutta/ddm_catalogs.pdf
- **Haimonti Dutta**, Hillol Kargupta, and Anupam Joshi, “Orthogonal Decision Trees for Resource-Constrained Physiological Data Stream Monitoring using Mobile Devices”, *High Performance Computing Conference (HiPC)*, pages 118 – 127, Goa, India, 2005.
URL: <http://www1.ccls.columbia.edu/~dutta/hipc05.pdf>
- Hillol Kargupta and **Haimonti Dutta**, “Orthogonal Decision Trees”, *The Fourth IEEE International Conference on Data Mining (ICDM)*, pages 487–490, Brighton, UK, 2004.
URL: <http://www1.ccls.columbia.edu/~dutta/ICDM04.pdf>
- Haimonti Dutta, Hillol Kargupta, Souptik Datta and Krishnamoorthy Siva Kumar, “Privacy Preserving Data Mining and Random Perturbations”, *Workshop on Privacy in the Electronic Society*

(in association with 10th ACM Conference on Computer and Communications Security),2003.
URL: <http://www1.ccls.columbia.edu/~dutta/WPES.pdf>

- Vasilis Megalooikonomou, **Haimonti Dutta**, Despina Kontos, “Fast and Effective Characterization of 3D Region Data”, *International Conference of Image Processing (ICIP)*, pages 421 – 424, Rochester, NY, 2002.
URL: <http://www1.ccls.columbia.edu/~dutta/ICIP2002.pdf>

TECHNICAL REPORT AND OTHER NON-REFEREED PUBLICATIONS

- **Haimonti Dutta**, Aayushee Gupta, Srikanta Bedathur, and Lipika Dey. “A Machine Learning Framework to Quantitative Prosopography”, arXiv:1801.10080. <https://arxiv.org/abs/1801.10080>, 2018.
- Haimonti Dutta and Ashwin Srinivasan, “Consensus-Based Modelling using Distributed Feature Construction”,<http://arxiv.org/abs/1409.3446>, 2014.
- Nipun Batra, Amarjeet Singh, Pushpendra Singh, Haimonti Dutta, Venkatesh Sarangan, Mani Srivastava, “Data Driven Energy Efficiency in Buildings”, <http://arxiv.org/abs/1404.7227>, 2014.
- Haimonti Dutta, William Chan, Deepak Shankargouda, Manoj Pooleery, Axinia Radeva, Kyle Rego, Boyi Xie, Rebecca J. Passonneau, Austin Lee, Barbara Taranto
“Leveraging Subjective Human Annotation for Clustering Historic Newspaper Articles.” CoRR abs/1208.3530, 2012.
URL: <http://arxiv.org/abs/1208.3530>
- Shen Wang and Haimonti Dutta, “PARABLE: A PARallel RAndom-partition Based Hierarchical Clustering Algorithm for the MapReduce Framework”, Technical Report, CCLS-11-04, 2011.
URL: <http://www1.ccls.columbia.edu/~dutta/CCLS-11-04.pdf>
- Phil Gross, Ansa Saleb-Aouissi, Haimonti Dutta and Albert Boulanger, “Susceptibility Ranking of Electrical Feeders: A Case Study”, Technical Report, CCLS-08-04, 2008.
- Cynthia Rudin, Becky Passonneau, Axinia Radeva, Haimonti Dutta, Nandini Bhardwaj, Jawwad Sultan. “Columbia/Con Edison Project on Secondary System Events, Phase 2 Final Report”, Center for Computational Learning Systems, Columbia University, December 2007b.
- Haimonti Dutta, “Empowering Scientific Discovery by Distributed Data Mining on the Grid Infrastructure”, A Proposal for Doctoral Research, submitted to the CSEE Department, UMBC, 2006.
URL: <http://www.csee.umbc.edu/~hdutta1/proposal.pdf>
- Madhu Nayakkankuppam and Haimonti Dutta, “Maximum Likelihood Phylogenetic Tree Construction”, Extended Abstract submitted to Graduate Research Conference, University of Maryland, Baltimore County, 2003.
- Vasilis Megalooikonomou, Marc Sobel, Haimonti Dutta , Despina Kontos , “A statistical approach on effective dimensionality reduction for efficient classification of spatial regions of interest”, 2002.

References

To be given on request.