Subhabrata Majumdar

Senior Inventive Scientist, Data Science and AI Research, AT&T Chief Data Office
e-mail: subho@research.att.com; zoom.subha@gmail.com web: shubhobm.github.io

Expertise

Trustworthy machine learning: fairness, transparency, privacy and robustness.
Predictive models on big data: machine learning and statistical techniques, NLP.
Statistical machine learning: high-dimensional models, graphical models, feature selection, hypothesis testing.
Tools: R, Python, C++, Hive/SQL, Hadoop.

Education

- PhD Statistics, University of Minnesota - Twin Cities, 2017. Advisor: Snigdhansu Chatterjee.
- M.S. Statistics, Indian Statistical Institute, 2012.
- B.S. Statistics, Indian Statistical Institute, 2010.

Professional experience

AT&T Data Science and AI Research, Aug 2018-present

- Part of AI governance core team for implementing responsible ML practices based on original research.
- Current research projects involve developing novel fairness, transparency and privacy-centric methods for business problems, in diverse application areas such as NLP, spatial data, and network analysis.
- Major past projects in other areas involve viewership-centric content analytics and ad intelligence methods.
- Produced 10+ research papers and 10+ patent filings in ~2.5 years through cross-team collaborations.
- Member of multiple (past and present) organizational committees and hiring committee.

IBM Research, May 2016-Aug 2016

- Research intern in the IBM Social Good fellowship program.
- Collaborated with scientists in the Data Science group, and the Cary Institute of Ecosystem Studies to mine biological and ecological data and devise cognitive algorithms that can determine which primates are carriers for the Zika virus.
- Work led to publication in the journal Epidemics.

Santander Consumer USA, May 2015-Aug 2015

- Worked with the Statistical Analysis team on implementing new machine learning methods in Loss Forecast Score prediction.
- Used random forest, XGBoost and bagging to achieve performance improvements over current model.

National Marrow Donor Program, June 2013-Aug 2013

- As a statistician intern, developed a spatial algorithm for data-driven marrow donor recruitment for Leukemia patients with rare alleles in 2013 summer.
- Collaborated with scientists at the bioinformatics division.

Leadership

- Driving the development and adoption of trustworthy machine learning methods in AT&T through technical leadership.
- Co-founder and co-organizer of the Trustworthy ML Initiative (https://www.trustworthyml.org).
- Managed a team of 4 data scientists to collaborate on a volunteer project with UNICEF officials in developing a machine learning platform for air-quality prediction (https://www.solveforgood.org/proj/41).

Publications

Notable works

- **Majumdar, S.** Fairness, Explainability, Privacy, and Robustness for Trustworthy Algorithmic Decision Making. In Big Data Analytics in Chemoinformatics and Bioinformatics, in press, 2021, published by Elsevier.

- *Flynn, C., **Majumdar, S.** and Mitra, R. Evaluating Fairness in the Presence of Spatial Autocorrelation, SDSS-2021, arXiv preprint arXiv:2101.01703, 2021. **Alphabetical authors*.
- Farias, V., Timbo, F., Flynn, C., Machado, J., **Majumdar, S.,** Srivastava, D. Local Dampening: Differential Privacy for Non-numeric Queries via Local Sensitivity, PVLDB, 14(4), 521-533, 2020.
- Rustamov, R. and **Majumdar, S**. Intrinsic Sliced Wasserstein Distances for Comparing Collections of Probability Distributions on Manifolds and Graphs, arXiv preprint arXiv:2010.15285, 2020.
- *Dodwell, E., Flynn, C., Krishnamurthy, B., **Majumdar, S.** and Mitra, R. System to Integrate Fairness Transparently: An Industry Approach, arXiv preprint arXiv:2006.06082, 2020. **Alphabetical authors*.
- Ghosh, A. and **Majumdar, S.** Ultrahigh-dimensional Robust and Efficient Sparse Regression using Non-Concave Penalized Density Power Divergence, IEEE Transactions on Information Theory, 66(12), 7812-7827, 2020.
- 7 authors. Computer-Assisted and Data Driven Approaches for Surveillance, Drug Discovery, and Vaccine Design for the Zika Virus, Pharmaceuticals, 12(4), 157, 2019.
- 11 authors. Confronting data sparsity to identify potential sources of Zika virus spillover infection among primates, Epidemics, 27, 59-65, 2019.
- **Majumdar, S.** and Basak, S. C. Beware of external validation!-A Comparative Study of Several Validation Techniques used in QSAR Modelling, Current Computer-Aided Drug Design, 14(4), 284-291, 2018.
- **Majumdar, S.** and Michailidis, G. Joint Estimation and Inference for Data Integration Problems based on Multiple Multi-layered Gaussian Graphical Models, arXiv preprint arXiv:1803.03348, 2018.
- 10 authors. Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning, KDD-2015, 2039-2047, 2015.

Please see my Google scholar for a full list.

Talks (invited/refereed)

06/2021 ASA 2021 Symposium on Data Science and Statistics.
05/2021 International Indian Statistical Association Conference-2021.
02/2021 ASA 2021 Conference on Statistical Practice.
01/2021 Plenary talk in All India Council for Technical Education Faculty Development Programme.
12/2020 Data Science Salon Virtual.
11/2020 Indian Institute of Technology, Kanpur Data Science Seminar Series (3 invited talks).
02/2020 Invited talk on Data Science Research in AT&T in 3rd NISS Virtual Industry Career Fair.
11/2019 3rd AT&T Graduate Student Symposium.
03/2019 NYC Women in Machine Learning & Data Science meetup.
05/2018 Savvysherpa, Inc., Minneapolis, MN.
05/2018 International Indian Statistical Association Conference-2018, Gainesville, FL.
12/2017 International Indian Statistical Association Conference-2017, Hyderabad, India.
12/2017 Indian Statistical Institute, Kolkata, India.
04/2017 ASA Twin Cities Chapter Spring Meeting, Minneapolis, MN.
08/2016 International Indian Statistical Association Conference-2016, Corvallis, OR.
07/2013 National Marrow Donor Program, Minneapolis, MN.

Academic experience

- Postdoctoral researcher, University of Florida Informatics Institute, July 2017 – Aug 2018.
- Graduate Assistant, University of Minnesota Twin Cities, Minneapolis, MN, Sep 2012 – May 2017.
- Data Science for Social Good fellow, University of Chicago, Chicago, IL. June 2014 – Aug 2014.

Major awards

University of Minnesota

- Martin-Buehler Award in Statistics 2016-2017, awarded by School of Statistics.
- Interdisciplinary Doctoral Fellowship 2016-2017, awarded by the Graduate School.
- Social Community Building Grant 2015-2016, awadred by Council of Graduate Students.

Conference awards

- Best Student Paper Award in theory and methods, 2016 International Indian Statistical Association conference, Corvallis, OR.
- School of Statistics Travel Awards, 2014-2016.
- 5th International Workshop on Climate Informatics travel award (funded by National Science Foundation (NSF)), 2015.