

Koustuv Saha

Curriculum Vitae

341A, Technology Square Research Building
85 Fifth Street NW, Atlanta, GA 30308, US
☎ (+1) 404 692 9496
✉ koustuv.saha@gatech.edu
🌐 www.koustuv.com

A computer science engineer and researcher interested in the interdisciplinary domain of computational social science and its applications in precision medicine, working with real-world problems from natural language processing, machine learning, data science, and statistical modeling. Current research involves social media sensing for mental health with a focus on situated communities (such as college campuses and workplaces).

Education

2016– **Doctor of Philosophy Student, Computer Science.**

Advisor: Dr. Munmun De Choudhury, Specialization: Social Computing
School of Interactive Computing, Georgia Institute of Technology

2008–2012 **Bachelor of Technology (Hons.), Computer Science and Engineering.**

Guide: Dr. Sujoy Ghose, B.Tech Project: Ontology based Semantic Information Retrieval
Computer Science and Engineering, Indian Institute of Technology Kharagpur

Publications

2018 **Saha, K.**, Weber, I., & De Choudhury, M. (2018). A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses. In Proceedings of the International AAAI Conference on Web and Social Media., *Presented at ICWSM'18*.

2018 **Saha, K.**, & De Choudhury, M. (2017). Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. In Proceedings of the ACM on Human-Computer Interaction, 1(CSCW, 92)., *To be presented at CSCW'18*.

2017 **Saha, K.**, Chan, L., de Barbaro, K., Abowd, G., & De Choudhury, M. (2017). Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(95)., *Presented at UbiComp'17*.

2017 **Saha, K.**, Weber, I., Birnbaum, M. L., & De Choudhury, M. (2017). Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates. Journal of medical Internet research, 19(5)., *Presented at Mental Health Symposium, CHI'17*.

2017 Sharma, E.*, **Saha, K.***, Ernala, S. K.*, Ghoshal, S.*, & De Choudhury, M. (2017). Analyzing Ideological Discourse on Social Media: A Case Study on the Abortion Discourse. In Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas (CSS), (* co-primary authors).

Patents

2015 Patel, L., **Saha, K.**, Das, S., & Biswas, A., A System and Method for Sentiment Analysis of Text., *US Provisional Patent Filed, Aug'14, US Patent Filed, Aug'15*.

2015 Patel, L., **Saha, K.**, Das, S., & Biswas, A., A System and Method for scalable and continuous learning and annotation of data., *US Provisional Patent Filed, Mar'15*.

2014 Bose, A., **Saha, K.**, Das, S., & Biswas, N., A System and Method for Predicting Compliance Issues., *US Provisional Patent Filed, Feb'14*.

Experience

- Aug 2016– **Graduate Research Assistant**, *Georgia Tech*, School of Interactive Computing.
PhD (Computer Science), specializing in Social Computing. Research interest primarily involves using social media as a passive sensor for characterizing and predicting mental wellbeing attributes of individuals. Broadly, motivated towards devising novel techniques of leveraging data for real-world problems.
- May 2018 – **Research Intern**, *Max Planck Institute for Software Systems (MPI-SWS)*.
July 2018 Worked on the problem of how social media based advertisement platforms enable advertisers to micro-target their audience in a malicious way. In particular, experimentally studied how Facebook ads from the Russian IRA run prior to 2016 US elections exploited Facebook's targeted advertising infrastructure to efficiently target ads on divisive or polarizing topics (eg, immigration, race-based policing) at vulnerable sub-populations.
- May 2017 – **Research Intern**, *Fred Hutch*, Hutch Data Commonwealth.
July 2017 Worked in biomedical data sciences and computational social science. Primary research problem included high dimensional machine learning using genomics data to predict the risk of Colorectal Cancer. Also, worked on quantifying the 'impact' of tweets to analyze the outreach of Cancer Research information and awareness.
- Dec 2012 – **Senior Researcher**, *Abzooba Inc.*, X, NLP & Data Sciences R&D.
Aug 2016 As the product development lead of XPresso, Abzooba's flagship product, improvised techniques, and implemented modules for extracting information in an end-to-end automated, scalable, and efficient fashion. Worked with predictive analytics based business problems from finance, retail, and healthcare sector. Instrumental in conceptualization and implementation of Continuous Learning System (CLS). Ranked 1st at a US based retail giant's Hackathon for Compliance Risk Index (CRIX). Role involved dealing with algorithms and techniques in NLP and Machine Learning, as well as technological advances in software development at production level.
- NLP problems such as aspect based sentiment analysis, emotion categorization, document classification, semantic search, trending topics detection, ontology building, and so on.
 - NLP techniques like Semantic Role Labeling, Dependency Parsing, Word-Vectors, Topic Modeling, LSA.
 - Data Sciences problems like predicting compliance risk, member expenses, distributor costs, etc.
- Sep 2013 – **Visiting Faculty Member**, *Praxis Business School*, Analytics.
Jul 2014 Conducted courses on Big Data Analytics and Retail Analytics. Subjects included Hadoop, Map Reduce, Unix and Programming fundamentals, and Natural Language Processing.
- Aug 2012 – **Junior Research Associate**, *Infosys Ltd.*, Infosys Labs, R&D.
Nov 2012 Conceptualized & implemented an information extraction tool from unstructured financial reports. Contributed in ideation of a semantic similarity based recommendation system of research projects.
- May 2011 – **Research Intern**, *Siemens Information Systems Ltd.*, Corporate Technologies R&D.
July 2011 Set up a Hadoop cluster of 30 computers, installed Mahout and Katta for running jobs. Developed a tool for accessing Lucene indexes to search and analyze their content based on parameters.

Academic Distinctions

- 2008 Qualified in Top 0.5% at IITJEE 2008 and Top 0.3% at AIEEE 2008.
2006 Awarded National Talent Search Examination (NTSE) Scholarship by NCERT, Govt of India.

Miscellaneous

- Skills and Interest Computational Social Science, Crisis Informatics, Social Computing, Mental Health, Campus Wellbeing, Natural Language, Causal Inference, Time Series, Statistical Modeling, Machine Learning, Data Sciences
- Reviewing JMIR, ACM CSCW, ACM CHI, IEEE TAF
- Programming Java, Python, R, C, Bash, HTML, CSS, Javascript
- Libraries Stanford CoreNLP, LIWC, NLTK, WordNet, Pandas, Scikit-Learn, Vowpal Wabbit
- Framework MySQL, MongoDB, BigQuery, Hadoop, Neo4J
- Languages English, Bengali, Hindi

Last Updated: 2018/09/30