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 software professional and researcher interested in the interdisciplinary domain of social computing and its applications in precision medicine, working with real-world problems from natural language processing, machine learning, statistical modeling, and information retrieval. Currently involved in using social media as an unobtrusive medium of passive sensing for predicting and characterizing mental well-being attributes among individuals.

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

2017 CSCW **Saha, K.**, & De Choudhury, M. (2017). Modeling Stress with Social Media Around Incidents of Gun Violence in College Campuses. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing, **Accepted**., *To be presented at CSCW'18*.

2017 IMWUT **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) **Forthcoming**., *To be presented at Ubicomp'17*.

2017 JMIR **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 CHI'17*.

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

May 2017– **Research Intern**, *Fred Hutch*, Hutch Data Commonwealth.
Worked in the domain of 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. Additionally, worked on quantifying the 'impact' of tweets for analyzing the outreach of Cancer Research and Treatment information.

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.

Dec 2012 – Aug 2016 **Senior Researcher**, *Abzooba Inc.*, X, NLP & Data Sciences R&D.

As the product development lead of XPresso, Abzooba's flagship product, involved in core research with algorithms and techniques in natural language processing and machine learning, as well as the latest technological advances in software development at production-level. Conceptualized heuristics, improvised techniques, architected design, and implemented modules for solving specialized information distillation problems in an end-to-end automated, scalable, and efficient fashion. Worked with problems such as, aspect based sentiment analysis, emotion categorization, document classification, resume analytics, trending topics detection, and so on. Additionally, briefly involved in the product development of Xpresso Commerce - a guided commerce platform, where (semantic) natural language search is one of the primary problems. Also worked with distinctive predictive analytics and statistical modeling problems in data sciences for rational business requirements in the financial, retail, and healthcare sector.

- Worked on NLP problems such as, hashtag processing, LSH-based vector similarity, Feature-Opinion based parser, ontology augmentation heuristics, n-grams for sequence learning
- Instrumental in conceptualization and implementation of Continuous Learning System (CLS)
- Worked with latest NLP techniques like Semantic Role Labeling, Dependency Parsing, linguistic rules and feature engineering for Machine Learning Models for expressions classifier
- Used word-vectors for addressing unsupervised NLP problems. Notable problems include Aspect Categorization, Semantic Information Retrieval, Document Classification, Ontology Builder etc.
- Obtained highest rank of 23 (700 teams) at Kaggle's Global Sentiment Analysis problem
- Developed data annotation and QA tools. Key role in innovation of TagIT
- Ideation and implementation of Ontology Builder for efficient semi-automated domain modeling
- Ranked 1st at US retail giant's Hackathon for CRIX, a patented solution for predicting compliance risk
- Involved briefly on Large Group Predictive Modelling to predict member-wise monthly costs
- Innovated DeaDRIX (Dealer Distributor Relative Index) for sales' ranking of Dealer wrt Distributor
- Involved in the Abzooba Academy as a Trainer to conduct company-wide learning sessions
- Recognized and awarded for distinctive contribution towards Abzooba's analytics solutions

Sep 2013 – Jul 2014 **Visiting Faculty Member**, *Praxis Business School*, Analytics.

Conducted courses on Big Data Analytics and Retail Analytics. Subjects included Hadoop, Map Reduce, Unix and Programming fundamentals, and Natural Language Processing.

Aug 2012 – Nov 2012 **Junior Research Associate**, *Infosys Ltd.*, Infosys Labs, R&D.

- 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 – July 2011 **Research Intern**, *Siemens Information Systems Ltd.*, Corporate Technologies R&D.

- 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

Technology

Programming	Java, Python, R, C, C++
Framework	MySQL, MongoDB, Hadoop, Neo4J
Skills	Product Development, Machine Learning, Natural Language Processing, Statistical Modeling

Academic Distinctions

2008	Qualified in Top 0.5% (400,000) at IITJEE 2008 and Top 0.3% (1,000,000) at AIEEE 2008
2006	Awarded National Talent Search Examination (NTSE) Scholarship by NCERT, Govt of India

Last Updated: 2017/08/23