Soumyadeep Chatterjee

Contact Information	609 Grand Fir Ave, Apt. 7 Sunnyvale, CA 94086	soumyachat@gmail.com Phone: 612-321-1212

Summary Worked on A/B tests for improving product quality. Developed scalable models in Hadoop for ranking and personalization in mobile search on large scale distributed systems. Familiar with Hadoop framework and experienced in Python, R, Hive, Pig, Java, C++. Experienced in developing statistical and machine learning models for prediction and data analysis.

Employment

Data Scientist July 2016 – present
Data Science, Quora
- Worked on improving moderation and reporting queues
- Developed new metrics for monitoring and improving quality of notifications sent to users

Research Scientist October 2015 – July 2016
Personalization Sciences, Yahoo!
- Developed machine learning models for serving image and news results in Yahoo! mobile search improving metrics by 50%
- Developed end to end pipeline in Hadoop for extracting data from user activity logs, training ranking model and deployment

Research Intern June 2014 – August 2014
Personalization Sciences, Yahoo! Labs
- Developed latent variable models for user profiling on Yahoo! home, news and sports pages

Research Assistant May 2010 – August 2015
Dept. of Computer Science & Engg., University of Minnesota, Twin Cities
- Developed new statistical methods for regularization in high dimensional regression problems
- Developed pipeline for using machine learning models on climate datasets for physical hypothesis generation
- Applied machine learning methods for improving prediction in climate forecasting problems, achieving a lift of 20% in predictive accuracy

Education

University of Minnesota, Minneapolis, MN Fall 2009 – September 2015

Ph.D. in Computer Science & Engineering
- Thesis: High Dimensional Statistical Models: Applications to Climate
- Advisor: Arindam Banerjee, Ph.D

Jadavpur University, Kolkata, India July 2005 – June 2009

B. E. in Electronics & Telecommunications Engineering

Skills
- **Languages:** Python, MySQL, MATLAB, C++, Hive, Pig, Java (basic), R
- **Distributed Platform:** Hadoop
- **Operating Systems:** Windows, Linux, Mac OS X

Publications

1. **S. Chatterjee**, S. Liess, A. Banerjee and V. Kumar, "Understanding Dominant Factors for Precipitation over the Great Lakes Region", Thirtieth AAAI Conference On Artificial Intelligence 2016.
2. A. Asiaee Taheri, **S. Chatterjee** and A. Banerjee, "Regularized Structured Estimation in High-Dimensions with Noisy Designs", SIAM Intl. Conference on Data Mining 2016.
3. **S. Chatterjee**, S. Chen and A. Banerjee, "Generalized Dantzig Selector: Application to the k-support norm", Advances in Neural Information Processing Systems (NIPS), 2014.
4. A.Goncalves, P. Das, **S. Chatterjee**, V. Sivakumar, F. J. Von Zuben and A. Banerjee, "Multi-task Sparse Structure Learning", International Conference on Information and Knowledge Management (CIKM), 2014.
5. H. Wang, F. Fazayeli, **S. Chatterjee** and A. Banerjee, "Gaussian Copula Precision Estimation with Missing Values", International Conference on Artificial Intelligence and Statistics (AISTATS), 2014.
6. **S. Chatterjee**, A. Banerjee, S. Chatterjee and A. Ganguly, "Mixture of Regression Models for Precipitation Prediction", The Second International Workshop on Climate Informatics (CI), 2012.
7. **S. Chatterjee**, K. Steinhaeuser, A. Banerjee, S. Chatterjee and A. Ganguly, "Sparse Group Lasso: Consistency and Climate Applications", SIAM Intl. Conference on Data Mining 2012 (**Best Student Paper Award**).
8. **S. Chatterjee**, K. Bhattacharjee and A. Konar, "A Simple and Robust Algorithm for Microarray Data Clustering Based on Gene Population-Variance Ratio Metric", Biotechnology Journal, Vol. 4, Issue 9, 2009.
9. K. Bhattacharjee, **S. Chatterjee** and A. Konar, "A Novel Clustering Method for Gene Microarray Data Using Intra-Cluster Distance and Variance", IEEE Intl. Advanced Computing Conference, 2009 (IACC-09).

Awards & Scholarships

- Best Student Paper Award at the SIAM International Conference on Data Mining (SDM), 2012.
- 3 year Graduate School Fellowship (2009-2012) from Univ. of Minnesota Twin Cities.
- Travel Fellowship for poster presented at The Second International Workshop on Climate Informatics 2012.
- Travel Award for paper presented at SDM 2012.
- 4-year J. C. Bose National Science Talent Search (JBNSTS) Senior Scholarship, 2005–2009.