Swupnil Kumar Sahai

Columbia Univeristy
Department of Statistics
Room 1005 School of Social Work
1255 Amsterdam Avenue
New York, NY 10027

Phone: (408) 406-6076
Email: swupnil@stat.columbia.edu
Homepage: http://www.swupnil.com/
LinkedIn: http://www.linkedin.com/swupnil
GitHub: http://www.github.com/swupnil

Education

COLUMBIA UNIVERSITY
   Ph.D. in Statistics, 2013-2017

UNIVERSITY OF CALIFORNIA, BERKELEY
   B.A. (Honors) in Applied Mathematics, Statistics, & Economics; 2009-2013
   Chinese Language & Culture Minor

Awards & Grants

2014: Natural Sciences & Data Sciences Interface Grant, Columbia IDSE.
2014: Provost's Diversity Fellowship, Columbia University.
2013-2018: Dean's Fellowship, Columbia University.
2009-2013: Regents' and Chancellor's Scholarship, UC Berkeley.

Projects

INDEPENDENT PROJECT
Apple Watch Sports Application, 2014-2015

   · Developing a social sports analytics application for Apple Watch to change the way
     recreational athletes view their games and matches
   · Coding all of the iOS and WatchOS front end in Swift using Xcode
   · To be released alongside Apple Watch OS2 in Fall 2015

DEPARTMENT OF STATISTICS, UC BERKELEY
U.S. Presidential Election Prediction, Fall 2012 [Link]

   · Wrote a paper describing the Bayesian methodology of using a Multinomial likelihood for
     polling data, an informative state-wise Dirichlet prior incorporating previous election results,
     and sampling from a posterior distribution generated by Monte Carlo chains
   · Predicted every state's 2012 election outcome correctly, with the exception of Florida

DEPARTMENT OF STATISTICS, UC BERKELEY
Kaggle Data Prediction Competition, Spring 2012

   · Utilized lasso, ridge, support vector machines, nearest neighbors, and regression trees to
     predict the brightness of images shown to patients from 10,000+ brain voxel predictors
   · Led team to second place in Spring 2012 Machine Learning course

Teaching

DEPARTMENT OF MATHEMATICS, UC BERKELEY
Undergraduate Student Instructor, Spring 2013

> · Lectured and wrote quizzes for six weekly recitations of Math 53 (Multivariable Calculus)

Research

INSTITUTE FOR DATA SCIENCES AND ENGINEERING, COLUMBIA UNIVERSITY
Birth of the Universe and Fate of the Earth, Andrew Gelman and David Schiminovich, 2014-2015

> · Leveraging PyStan to analyze a 40TB data set consisting of 10^{12} photons recorded by the GALEX ultraviolet space telescope over a period of 10^8 seconds
> · Developing a scalable algorithm for posterior inference of hierarchical Bayesian models via parallel expectation propagation
> · Utilizing this algorithm to fit a hierarchical model to the relationship between infrared and ultraviolet radiation across different longitudinal regions of the observable universe, and detect outliers that may be indicative of radiation from the Big Bang

DEPARTMENT OF STATISTICS, COLUMBIA UNIVERSITY
Hard to Reach Populations & Secret Keeping, Andrew Gelman and Tian Zheng, 2014-2015

> · Utilizing an over-dispersed Poisson model in RStan to analyze aggregate relational survey data consisting of questions of the form "How many X's do you know?"
> · Estimating national lifetime prevalence of abortions and miscarriages from questions about how many women the survey respondents know who've had abortion/miscarriage
> · Calculating the extent of secret keeping by comparing national cancer prevalences to respondents' knowledge of lung, prostate & cervical cancer diagnoses within their networks
> · Presented findings at the 2015 Annual Eastern Sociological Society Conference in NYC

Employment

TESLA MOTORS INC.; PALO ALTO, CA
Data Science Intern, Summer 2015
Worked on the Reliability Engineering team to build a Python module facilitating Bayesian inference and visualization of fleet failure data, for use by engineers.

> · Built a holistic Python module for Bayesian fitting of 3-parameter Weibull distributions to censored failure data (i.e. data with car components that have failed and that have yet to fail)
> · Created a novel visualization to intuitively display a fitted Weibull distribution on top of failed and censored data histograms
> · Developed libraries to automatically generate animated visualizations of probabilistic simulations of an entire fleet of components failing over time, given current fleet data
> · Designed a novel equation with inverse logistic functions to optimally allocate reliability goals for future vehicle component systems

NAMELY INC.; BROOKLYN, NY
Data Science Intern, Spring 2015
Worked as a part-time Spring intern on the Data & Analytics team to build a novel salary recommendation engine as a standalone Namely product.

· Designed the statistical model for Namely's salary recommendation engine from scratch
· Built the model in Python using modules such as pandas, numpy, scipy, and scikit-learn
· Created an interactive front end with drop down menus to select job title, seniority, number of people managed, and geographic location, and subsequently recommend a salary range

ALLSTATE RESEARCH & PLANNING CENTER; MENLO PARK, CA
Predictive Modeling Intern, Summer 2013
Worked on the Analytics and Insights team to understand the partnerships between car insurance agents and life insurance specialists at Allstate, as recorded by a 150+ MB data set.

· Leveraged R and Tableau to conduct exploratory analysis of partnership network data and create novel metrics such as partnership density/richness & policy momentum/volitility
· Discovered that suboptimal combinations of these novel metrics over 6-12 month periods provide strong predictors of future network failure
· Presented findings to 60+ Research Center employees in a brown bag
· Won first place in Allstate's nation-wide intern social media competition

ILLUMINATE CONSULTING GROUP; SAN CARLOS, CALIFORNIA
Quantitative Analyst Intern, Summer 2012
Conducted statistical analyses of over 5,000 international students' performance across and within four major Canadian and Australian universities to make informed policy recommendations.

· Utilized Tableau to analyze post-secondary student performance measures
· Developed a pattern recognition technique to detect possible English testing fraud
· Wrote statistical meaning and policy implications into four individual institutional reports and one comprehensive benchmark report evaluating performance of various nationalities

APPLE INC.; MIAMI, FLORIDA
Finance Intern, Summer 2012
Reported to Senior Finance Manager of Apple's Latin America & Caribbean (ALAC) and was responsible for data collection, analysis, and quarterly reporting of ALAC financial performance.

· Managed and completely restructured the database housing all ALAC financial transactions to make financial report generation more efficient
· Coded a script to automatically convert new ALAC financial transactions into FileMaker format for integration into Cupertino's master database

Popular Press

"Triple Majors Make 27 at Cal Noteworthy." *San Francisco Chronicle.* January 2013. [Link]

Leadership

COLUMBIA DATA SCIENCE SOCIETY, COLUMBIA UNIVERSITY
Executive Board Member, 2014-2015

· Organized events with industry leaders in Data Science
· Led a public tutorial on Tableau in November 2014 (70 attendees) and February 2015 (50 attendees)

UNDERGRADUATE ECONOMICS ASSOCIATION, UC BERKELEY
President, 2012-2013

· Created Facebook page and weekly newsletter, and initiated mentorship program to foster academic and personal relationships between officers and members
· Organized and publicized a public lecture event for over 300 attendees featuring 2009 Economics Nobel Laureate, Dr. Oliver Williamson

Software Skills, Foreign Languages, and Interests

· Proficient in R, Python, Swift, Xcode, LATEX, and Tableau
· Comfortable with JAVA, C, and UNIX
· 8 Years of experience in reading/writing/speaking Mandarin Chinese; Fluent in Hindi
· Enjoy reading, volunteering, traveling, Poker, swimming, skiing, and tennis

Last updated: July 20, 2015