SIDDHANTH SABHARWAL

siddhanth.sabharwal1995@gmail.com - +1 (510) 585-6636 - 42701 Loma Drive, Fremont, CA 94539

EDUCATION

Columbia University in the City of New York, NY **2018 – 2019**
Master of Arts in Statistics
Overall GPA: 3.967/4.000

- **Masters Level Statistics Coursework**

Probability, Statistical Inference, Linear Regression Models, Statistical Computing, Elementary Stochastic Processes, Survival Analysis, Statistical Machine Learning, Advanced Machine Learning, Statistical Inference/Time-Series Modelling, Advanced Data Analysis

University of California - Davis, CA **2013 - 2017**
Bachelor of Science in Statistics and Computer Science
Overall GPA: 3.454/4.000, Statistics GPA: 3.778/4.000, Computer Science GPA: 3.330/4.000

- **Upper Division Statistics and Economics Coursework**

Analysis of Variance, Regression Analysis, Mathematical Statistics and Probability I/II, Multivariate Data Analysis, Analysis of Categorical Data, Nonparametric Statistics, Applied Time Series Analysis, Statistical Data Science Intermediate Macroeconomics Theory, Intermediate Microeconomics Theory, Game Theory, Financial Economics, Econometrics

Upper Division Computer Science and Math Coursework
Computer Architecture I, Computer Architecture II, Computer Graphics, Algorithm Design, Theory of Computation, Database Systems, Operating Systems, Information Interfaces (UI Design), Programming Languages, Ethics in an Age of Technology, Optimization, Probability

HONORS

- **Columbia University, December 2019:** Chair's List of Academic Achievement award
- **UC Davis, Fall Quarter 2017:** Dean's Honors List

RESEARCH EXPERIENCE

Columbia University, *Data Science Institute* **New York, NY**
Graduate Research Associate for Prof. Siddhartha Dalal **2019 - Present**

- Contributing to research project aimed at developing a graph classification model and techniques to find which nodes are causing the graph to be classified as fraudulent in the **Bitcoin** network
- Web scraped Bitcoin forums to find known fraudulent addresses
- Extracted shadow networks around fraudulent addresses, actors and their transactions using recursive techniques, **BlockSci** and **SNAP**
- Developing strategies to compare networks of varying number of nodes including zero-padding and simulation
- Trying various inductive learning algorithms such as **GraphSAGE** on graph data to generate embeddings
- Using the model trained on the graph embedding to identify other suspicious behavior in the Bitcoin network

Columbia University, *Data Science Institute* **New York, NY**
Graduate Research Associate for Prof. Kriste Krstovski and Prof. Yao Lu **2019 - Present**

- Contributing to research project to develop an index that ranks companies based on the amount of gender and race related discrimination encountered by their employees
- Web scraped company reviews from **Glassdoor** using **Selenium** and **BeautifulSoup**; Combining Glassdoor data with web scraped data from the **Equal Employment Opportunity Commission (EEOC)** to see if review data prior to the filing of the complaint could have predicted the complaint
- Detected occurrence of race and gender related language using **Latent Dirichlet Allocation topic modeling**
- Inferring race and gender and categorizing job shifts as promotions, demotions, and lateral moves in **LinkedIn** data

University of California Davis, *Visualization & Interface Design Innovation Lab (*vidi.cs.ucdavis.edu) **Davis, CA**
Undergraduate Researcher for Prof. Kwan-Liu Ma **2016 - 2017**

- Created dashboard and dynamic visualizations for supercomputer network performance using **JavaScript**
- Designed novel methodologies to extract insights from time series data using **R**
- Worked as a statistician to design VARIMA models for supercomputer processor behavior

PROFESSIONAL EXPERIENCE

Equinix **Redwood City, CA**
Senior Data Scientist, Marketing Strategy and Intelligence (MSI) **2015 - Present**

- Equinix is a company that provides data center space where other companies physically host their cloud deployments
- MSI is the team within Equinix that designs algorithms to rank prospects so our sales teams can go after the best targets
- Producing **revenue forecast models** and **wallet sizing models** using **random forest and boosting methodologies.** Improved our ability to size a prospect in terms of revenue and determine market penetration
- Designing **churn models** using **time-series and logistic regression methodology** to predict retention and loss amongst current customer pool
- Integrating the **prospect targeting program** with inputs in Java and VBA to find geographical locations with high revenue opportunities but no sales coverage
- Coding a **customer upselling tool** for internal marketing use that allowed upselling additional connectivity options to existing customers to increase revenue

LotaData **Berkeley, CA**
Data Analyst **2015 - 2016**

- LotaData sold maps that had listings of community events so companies could advertise at local events
- Designed interactive hyper local visuals for advertising purposes using **ElasticSearch, Kibana, and D3**
- Responsible for updating website with interactive visuals using **Node.js and Grunt.js**

eSilicon **San Jose, CA**
Technical Accounting/Finance Intern **2013 - 2014**

- eSilicon is semiconductor company that designs ASIIC components for other companies
- **Generated $3M of new business in a year** through designing a system that allowed users to configure a chip design online
- Designed Excel Macros to **automate the quote process** and **sub ledger and cost accounting tools** for sales marketing and finance teams

RESEARCH INTERESTS

- **Computational Statistics -** Variable Selection in High-Dimensional Data, Causal Inference, Analysis of Networks
- **Nonparametric Statistics**
- **Time Series Analysis**
- **Applications of Statistics in Marketing and Professional Sports**

PROGRAMMING LANGUAGES and SELECTED PACKAGES

- **R:** dplyr, plyr, ggplot2, glmnet, gbm, garet, e1071, iGraph
- **Python:** SNAP, BlockSci, PySpark, Pandas, Numpy, Selenium, BeautifulSoup, Multiprocessing
- **JavaScript:** D3
- **C, C++, Java, VBA**

SELECTED CLASS PROJECTS

- **Columbia University**
  - **Linear Regression Models:** Are the average male wages statistically different across race classes?
  - **Survival Analysis:** Do different treatments affect autoimmune liver disease survival?
  - **Advanced Machine Learning (In Progress):** Can we effectively transfer artist style for new creations?
  - **Advanced Data Analysis (In Progress):** What can an employer do to raise their employee ratings?
- **UC Davis**
  - **Regression Analysis:** Can body weight percentage be inferred via outward body measurements?
  - **Analysis of Categorical Data:** Can depression be diagnosed via various patient health scores?
  - **Nonparametric Statistics:** Does the length of cuckoo eggs determine on the host species' breed?
  - **Time Series Analysis:** Can the number of armed robberies in Boston, in the future, be predicted?
  - **Statistical Data Science:** Can a persons' gender be determined from their tweets (Twitter)?