Samuel Sharpe

223 West 115th Sreet, New York, NY 10026

sbs2193@columbia.edu ssharpe42.github.io ssharpe42 (301) 412-1930

Education

Columbia University **New York, NY**
M.S. in Computer Science, GPA: 3.86 *May 2020*
- Machine learning specialization

Columbia University **New York, NY**
B.S. in Operations Research, GPA: 3.98, Magna Cum Laude *Sep. 2013 - May 2015*
- *Stephen D. Guarino Memorial Award in Industrial Engineering* - Awarded to a senior with outstanding academic achievement and potential.

Oberlin College **Oberlin, OH**
B.S. in Mathematics, Economics Minor, GPA: 3.63 *Sep. 2010 - May 2013*

Skills

Programming/Technology
Python, R/RShiny, MATLAB, SQL, Airflow, Docker; Exposure to Kubeflow, Java, MongoDB, AMPL

Quantitative
Machine Learning, Optimization, NLP, Deep Learning, Stochastic Modeling, Speech Processing

Research & Publications

Publications
- *Visual natural language query auto-completion for estimating instance probabilities*.
  Samuel Sharpe, Jin Yan, Fan Wu, Iddo Drori; CVPR Language and Vision Workshop, 2019.

Course Research
- *Variational Autoencoders with Normalizing Flows for Unsupervised Anomaly Detection*. Spring 2020.
- *Stylometry in the Modern Era: Coreference and Voice for Authorship Attribution*. Spring 2019.

SiteRx *Jan. 2019 – May. 2019*
- Business to business start-up that connects patients with clinical trial sites through their physician.
- Research focuses on natural language processing of medical record text for deidentification.

Experience

Major League Baseball Advanced Media (MLBAM) **New York, NY**
Data Scientist *June 2018 - Present*
- Created pitcher-batter matchup projections and deployed in Kubernetes for use as prop betting lines in an MLB gaming app.
- Developed pitch classification neural network in Tensorflow and deployed endpoints for live pitch classification locally at ballparks.
- Designed pitch arsenal identification system using unsupervised learning to detect when pitchers add new pitches to their arsenal.

Booz Allen Hamilton **Washington, DC**
Lead Data Scientist *Jul. 2018 - Sep. 2018*
- Oversaw development of a resource reallocation model to optimize Immigration Judge (IJ) caseloads and shorten case times.
- Implemented a PoC IJ scheduling optimization designed to reduce rescheduled hearings for the Department of Justice (DOJ).

Staff Data Scientist / Modeling & Analysis Team Lead *Jan. 2017 - Jun. 2018*
- Formulated and implemented a multi-objective optimization model in MATLAB that presents leadership with tradeoffs between operational and budgetary objectives through sets of pareto-optimal resource allocations.
- Model endorsed by Department of Homeland Security (DHS) and awarded a DHS funded Winter Study.

Data Scientist *Sep. 2015 - Dec. 2016*
- Led implementation of probabilistic forecasts to inform risk-oriented budgeting that prevented over-appropriation.
- Utilized NLP, clustering, queuing simulation, and other techniques in R and Python on ad-hoc projects influencing client strategy.

Mindshare **New York, NY**
Data Analytics Intern *Jun. 2014 - May 2015*
- Developed attribution models to influence media spend and visualized twitter word-brand associations with NLP techniques.