

Caleb E. Strait

Data Scientist

(413) 884 - 4767
caleb.strait@gmail.com
Greater NYC Area

calebstrait.com
github.com/calebstrait
linkedin.com/in/calebstrait

Education

University of Rochester

Ph.D., Brain & Cognitive Sciences, May 2016
M.A., Brain & Cognitive Sciences, January 2014
Computational Neuroscience of Decision Making

Oberlin College

B.A., Psychology, May 2011
Minor, Computer Science
Concentration, Cognitive Science

Skills

Machine Learning: K-Means, PCA, SVM, Linear & Logistic Regression, DNN, CNN, RNN, Decision Trees, Recommendation Systems, Natural Language Processing
Data Analysis: Data Visualization, Statistical Modeling, Insight Mining
Python: Flask, Pandas, Numpy, Jupyter Notebook, SciKit-Learn, Matplotlib
GCP, AWS, Github, Java, C/C++/C#, Matlab, R, HTML, CSS, JS, API, SQL

Experience

Senior Data Science Consultant, Virtusa, March 2018 - Present

- *Partner Delivery Manager*, Google Cloud Consulting, Google
 - Worked onsite at Google NYC with ML engineers from across the country, consulting for client companies, analyzing data, and evaluating machine learning use cases for potential implementation via Google Cloud Platform. Focus topics: natural language processing and recommendation systems. (Python, GCP)
- *Fellow*, Advanced Solutions Lab, Google
 - Implemented a cloud-based financial transaction fraud detection system: data ingress via Google BigQuery, feature engineering with Google Datalab, classification via a Deep Neural Network estimator in TensorFlow, model training and evaluation on Google Cloud ML Engine, and user interface via Flask on Google App Engine. (Python, GCP)

Data Science Fellow, Insight Data Science, Jan 2017 - Dec 2017

- Devised and wrote custom recommendation algorithms for Insight client companies.
- Built a proof-of-concept web application around one of these algorithms. "Nextgame" gives increasingly personalized video game recommendations by prompting user feedback for each of a series of recommended games. Wrote a back end that collects data from web scraping and APIs, cleans data with regular expressions, and stores data in a PostgreSQL database hosted with AWS. The app averages reviewer scores to create an initial item recommendation hierarchy, then uses reinforcement learning to recalculate that hierarchy as the user gives feedback on recommendations. (Python)
- Analytically validated the app's performance showing that there were significantly more users than we would expect by chance for whom recommendations were more positively received after incorporating more feedback. (Python)
- Presented the algorithm and web app to over a dozen Insight client companies, including Facebook, Amazon, and Microsoft. (PowerPoint)

Computational Neuroscientist, University of Rochester, Aug 2011 - June 2016

- Estimated mid-decision neuronal spike frequencies from extracellular voltage measurement time series data using PCA and K-Means. (Matlab)
- Wrote a series of custom signal processing toolkits for use with decision-making experiment datasets whose formats I had standardized. Used sliding-window logistic regression models of neuronal spike frequencies to characterize how the recorded brain area's neurons encode decision parameters. (Matlab)
- Published these findings across four 1st-author peer-reviewed articles in high impact-factor neuroscience journals, such as *Neuron* and *PLOS Biology*.