Matar Haller

408.431.7640
matarhaller@gmail.com
github.com/matarhaller
linkedin.com/in/matarhaller
San Francisco, CA

Tools

Proficient: Python, MATLAB, SQL, git
Familiar: Scala, Apache Hadoop ecosystem (Kafka, Spark, Kudu, Impala), R, Lisp, Tableau, bash, HTML

Experience

Data Scientist, Winton Capital, San Francisco, CA **2016-present**

- Constructed real-time inflation indices to be used in trading algorithms by tracking hundreds of thousands of online prices.
- Collaborated to engineer infrastructure to read and parse streams of scraped html data from Kafka into Kudu using Spark.
- Integrated Jupyter notebooks with Impala to provide simple access to rapidly changing distributed data. Researchers in Zurich, London, and Oxford utilize the data and analyses developed for the current project in their own research.

Data Science Fellow, Insight Data Science, Palo Alto, CA **2016**

- Consulted for startup DeepGram to automatically segment unidentified speakers in recorded audio files.
- Engineered unsupervised algorithm combining spectral decomposition, PCA, and hierarchical clustering (matarhaller.com)
- Built novel validation dataset with known ground truth by scraping and splicing audiobooks (72% accuracy).
- Delivered code for use in existing predictive models and company proof of concept pitches.
- Accepted as talk to Strata-Hadoop 2017 in San Jose, CA.

Graduate Student Researcher, Helen Wills Neuroscience Institute, UC Berkeley **2010 – 2016**

- Recorded activity from surgically implanted electrodes on human brains. Implemented time series analyses, PCA, clustering, ridge regression and permutation tests to link neuronal activity with behavior. In review at *Nature Human Behavior.*
- Developed method for extracting oscillatory components from power spectra using linear regression, k-means clustering, curve fitting and cross validation. Precision and recall match human performance.
- Supervised and mentored research assistants and graduate students. Taught experimental design, electrophysiology recording, and analysis pipeline using Python and MATLAB.

Graduate Student Instructor, UC Berkeley **2005 – 2013**

- Collaborated with graduate students and professors to construct lesson plans for Neuropsychology, Drugs & the Brain, and Cognitive Science courses, each with 100-600 students each.
- Awarded Outstanding Graduate Student Instructor. Average evaluation: 6.7/7

Projects

OkNLP **2016**

- Combined NLP (tf-idf, tokenization) with machine learning (non-negative matrix factorization, grid search, cross validation) to analyze free text and demographic information from online dating profiles.
- Identified unintentional signals of drug usage status in free text self-descriptions across 60k user profiles.
- Presented as talk at SciPy 2016 in Austin, Texas and published in conference proceedings.

Software Carpentry & Women in Tech **2014**

- Taught scientific computing skills including Python, shell, and version control to Stanford graduate students.
- Taught programming using Jupyter notebooks to high school girls at PyData Silicon Valley.

Military

Tank Instructor, Armored Corps, Israel Defense Force **2007 – 2009**
Trained recruits, commanders and officers on the weapons control system of the Merkava Mark II tank.

Education

Ph.D., Neuroscience, UC Berkeley **2010 – 2016**
National Science Foundation Graduate Research Fellow
B.A., Cognitive Science; Psychology, UC Berkeley **2003 – 2007**
Highest Distinction in General Scholarship, Departmental Citation in Cognitive Science