Daniel Breen

619.558.8068 • dnlbreen@gmail.com
http://danielbreen.net/projects/

Physics PhD grad excited by using data mining tools to derive insights from real world data. 3 years experience creating knowledge and using python and data analytics to deliver results to stakeholders in the biomedical and bioengineering research community.

Experience

UC San Diego   **La Jolla, CA**
Graduate Research Assistant   *August 2014– 2017*

- Developed innovative cross disciplinary methods for stakeholders in two major projects in primary roles. Work resulted in a published paper, an invitation to publish in a selective journal, an oral presentation at SIAM DS17, and posters at SfN, MBI, and bioCAS 2016.

- Wrote python scripts and ran SGE parallel batch jobs on ROCKS cluster to process neural data and engineer features. Analyzed 3000x100 tabular data with pandas, and used scikit-learn's decision trees and random forest variable importances to identify differences between diseased and healthy neurons consistent with clinical observations of Alzheimer's patients.

- Established a mapping between configurable and true parameter values on neuromorphic VLSI chip, leading to emulation of a biological neuron on VLSI hardware for the first time.

- Extracurricular work: Wrote python scripts to scrape and clean recipes from foodnetwork.com, discovered ingredients characterizing ethnicity of cuisines using wordclouds, gensim's lda topic modeling, and deployed app online using flask and heroku.

- Analyzed Zillow's home sale prices, Wikipedia page on college towns, and GDP data with pandas and scipy to address question of whether housing prices in college towns are less affected by recessions. Published jupyter notebook lead to invitation to be a mentor for a new Coursera data science course.

Education

UC San Diego   **La Jolla, CA**
PhD in Physics, GPA:3.6   *2011–2017*

UC San Diego   **La Jolla, CA**
MS in Physics, GPA:3.6   *2011–2013*

New Mexico Tech   **Socorro, NM**
BS in Physics   *2007–2011*

Technical Skills

- **Machine Learning:** random forests, decision trees, neural networks, feature engineering

- **Statistical Methods:** linear and logistic regression, hypothesis testing, confidence intervals, bootstrap

- **Programming:** python (pandas, scikit-learn, matplotlib, numpy, scipy, gensim), linux, R, SQL, git

- **Mathematics:** differential equations, linear algebra, probability, statistics, time series