JOSIAH DAVIS

josiahjdavis.com | 415-500-5958 | josiah.j.davis@gmail.com

M.A., Statistics – University of California, Berkeley

August 2016 – May 2017

- Team lead for course projects in causal inference, optimization and statistical programming.

B.S., Mechanical Engineering – University of Maryland

August 2006 – Spring 2010

- Senior design team project chosen to be primary example for school's Mechanical Engineering design textbook.

Slalom Consulting, Lead Data Scientist — San Francisco, CA

May 2015 – Present

Lead Data Scientist for San Francisco office, directing technical aspects of data science projects, presenting regularly to senior-clients, mentoring junior data scientists, and providing expert contributions to business development efforts.

- Forecasted hourly customer behavior using multi-level machine learning (R – earth, rpart, tidyr, purrr).
- Analyzed customer bias in text of Yelp reviews (Python – NLTK, scikit-learn; R – tm, stringr, openNLP, syuzhet, plyr).
- Measured concentration in workload/asset distribution with the Gini coefficient (Python – pandas, numpy; Tableau).
- Presented at "Enterprise Applications of the R Language" Conference in Boston on Natural Language Processing in R.

Deloitte Consulting, Data Scientist — Washington, D.C.

February 2012 – May 2015

Data Scientist for Federal Government clients with experiences in machine learning, metric design and hypothesis testing.

- Created tree-based machine learning models to predict the probability of rework (R – randomForest, rpart).
- Derived and created a new estimate of latent process complexity (Python – pandas).
- Conducted a program evaluation of a multi-billion technology investment using survival analysis (R – survival).
- Earned the outstanding performance award two times for client work.

General Assembly, Data Science Instructor — Washington, D.C.

October 2014 – May 2015

Co-instructor for two iterations of the 66-hour course on Data Science covering the data science pipeline with a focus on supervised and unsupervised machine learning (Python – scikit-learn, pandas, numpy, matplotlib, statsmodels).

Causal Inference

- Directed Acyclic Graphs
- Potential Outcomes
- Backdoor Criteria
- G-computation formula
- Super Learning

Machine Learning

- Clustering
- Decision Trees
- Random Forests
- Ensemble Learning
- Cross-validation

Statistics

- Linear Modeling
- Model Checking
- Regularization
- General Linear Modeling
- Hypothesis Testing

Programming

- Python - pandas/numpy
- Python - scikit-learn
- R - dplyr/tidyr/ggplot2
- R - rpart/randomForest
- R - devtools/roxygen2