AVA HOFFMAN

Data Scientist · Ecologist

avamariehoffman@gmail.com
(804) 687-7476
/in/avahoffman
avahoffman
www.avahoffman.com

EXPERIENCE

Data Scientist
The Boston Consulting Group - Gamma
Boston, MA · March 2019+

- Built PySpark productionalized nested logistic regression model (>85% accuracy) for assessing risk in commercial customers of a Fortune 500 Financial Institution
- Engineered daily data insights piipeline to provide recommended products for commercial bank customers
- Implemented repository logging and error handling for an automotive manufacturer's inventory intelligence software
- Lead theory and methodology workshops for client teams

Data Science Fellow
Insight Data Science
Remote Program · September 2018 - March 2019

- Launched National Perks Project to improve visitor experience in crowded National Parks
- Leveraged large National Park Service, NOAA, & web-scraped datasets to forecast crowds (FB Prophet and ARIMA models) & weighted user preferences with euclidean distance for personalized experience
- Built a customized web app using Git, Python, Flask, & Heroku providing visitors with an actionable recommendation & access to further resources

USDA NIFA Predoctoral Fellow
Colorado State University
Fort Collins, CO · January 2017 - December 2018

- Designed Blue Grama Diversity Project to inform natural area stakeholders about genetic diversity in a foundational prairie grass, blue grama, leading a 6-member team
- Pioneered genomic feature detection (sequencing) of a key grass species, discovering >9,000 genomic features to cluster populations & guide conservation from 15GB of data
- Modeled hierarchical linkages between genomics, populations, & plant appearance, developed analytical workflows using R and RStan, & communicated genetic clusters in ggplot+shiny app

Doctoral Researcher
Colorado State University
Fort Collins, CO · August 2013 - December 2016

- Discovered diversity within key prairie grasses in response to drought to guide management of grassland resources
- Optimized analytics pipeline for genes & contrasted fluctuation of >100,000 genes under different conditions using bash scripts
- Quantified predictive relationships among plant, traits, & ecosystem by implementing trait feature reduction, Bayesian hierarchical models, path analysis, module clustering, literature data mining, linear regression, & meta-analysis in R, leading to domain innovation

EDUCATION

PhD · Ecology · Colorado State University
Fort Collins, CO · May 2019

BS · Biology · University of Virginia
Charlottesville, VA · May 2012

TOOLS & PACKAGES

Python

scikit-learn · pandas · NumPy · SciPy · pytest · statsmodels · seaborn · matplotlib

R

RStan · shiny · leaps · lavaan · dplyr · reshape2 · bioconductor · sva · vegan · ggplot2 · rmarkdown · RStudio

SQL · PySpark · git · Unix · Tableau · Alteryx · Markdown · LaTeX · SAS

TECHNIQUES

statistics

Bayesian methods · hierarchical models · mixture models · multivariate statistics · ANOVA · MANOVA · PERMANOVA · dimensionality reduction · repeated measures

supervised learning

random forest · time series · linear, nonlinear, logistic regression · LDA · SEM

unsupervised learning

PCA · k-means · feature engineering

ACHIEVEMENTS

Sustainability Leadership Fellow
Colorado State University · 2017-2018

Vice President for Research Fellow
Colorado State University · 2016-2017

14 peer-reviewed publications
Google Scholar · 2010-2019