Casson Stallings, Ph.D.

[cassonstallings@gmail.com](mailto:cassonstallings@gmail.com) - 919.264.3692

www.linkedin.com/in/cassonstallings

github.com/CassonStallings/Publications

Career Summary

Data scientist with more than 30 years of proven experience supporting science with quantitative analysis and software development. Extensive and diverse experience gathering, combining, and analyzing data. A successful history consulting with clients in the government, business, and academia.

Highlights

* Technical lead on numerous projects for a government contractor; provided technical direction and facilitated communication and collaboration between clients, partners, and contractors.
* Applied feature-based molecular networking to LC/MS metabolomics data, compared it to internal curation and presented results to research staff for consideration.
* Processed financial transaction data for 4 million people. Designed and constructed an AWS Redshift data base, and looked at receipts, customer acquisition, and growth trends of private companies. Determined basic facts, produced visualizations, and then worked with marketing to generate publications in the financial media.
* Developed methodology to increase the classification accuracy of satellite imagery by combining geo-referenced vector data and maximum likelihood spectral classification.

Experience

Data Scientist to Principal Data Scientist / Metabolon **2017 to 2023**

* Designed and coded the framework and machine learning components for software identifying compounds in untargeted LC/MS metabolomics data. This reduced manual processing time for projects from a week or more to as little as an hour removing a major constraint to the company’s growth.
* Analyzed the variability in human curation to better evaluate the automated processing of metabolomics data. Provided ad-hoc database queries and analyses to support the business and scientists. Supported, maintained, documented, and tested software developed by the Data Science team.

Senior Consultant-Data Science / Mammoth Data **2015 to 2016**

* Developed complex Tableau reports for multiple departments at a mortgage lender. The process included working with the company analysts to understand current reports generated primarily by hand in spreadsheets, adapting analysis to methods appropriate for Hadoop-based data, implementing Impala queries, some over 1000 lines of SQL and pulling data from 10+ data sources each.
* Assisted in automating ETL into the data lake, primarily using Kettle and a series of shell and Hive scripts, to incorporate database tables and spreadsheets into Hive/Impala.
* Mentored company analysts in use of Tableau, the Hadoop cluster, Kettle, Hive, and Impala.
* Analyzed data from Moogfest in Spark for marketing purposes.

Data Scientist, FutureAdvisor **2014 to 2015**

* Identified significant weaknesses in free customer diversification compared to managed portfolios which through customer education could be used to convert free to premium customers.
* Generated data stories from a large financial transaction database for marketing purposes.

Lead Mathematician / Statistician, **Alion Science and Technology** **2005 to 2015**

* Carried out predictive machine learning and spatial analysis to estimate human exposure to pollutants and identify at-risk populations. Assisted clients in comparing this with health data from Duke Hospital.
* Designed and coded a multi-pathway Monte Carlo model predicting human exposure to pollutants. Written in SAS, the model combines diverse data sources and generates billions of records for analysis.
* Statistically characterized the stability of pollutant samples, giving researchers much greater flexibility in collecting and storing samples.

Senior Systems Analyst, ManTech Environmental Technology 1998 to 2005

* Designed and coded, from scratch, the web interface for OasisGIS using JavaScript and Pascal.
* Updated the MS Access version of the Consolidated Human Activity Database (CHAD), incorporating flexible query, stochastic sampling, new study data, and spatial-temporal links to weather data.

Education

* Ph.D. in Forestry with a minor in Water Resources, North Carolina State University, Raleigh NC

Research Focus: Accuracy improvement and error propagation when combining satellite imagery, GIS, and complex simulation models

* M.S. in Biological Sciences, University of Southern California, Los Angeles CA

Research Focus: Development of predictive models based on satellite imagery of the oceans

* B.S. in Biology, Portland State University, Portland OR

Skills

Techniques: Machine learning, data mining, modeling, exploratory and spatial analysis, statistics

Programming: Python and common libraries, Dask, Azure DevOps (for Git, CI, and planning)

Databases: SQL, PostgreSQL, Oracle, MongoDB

Communication: Journal articles, reports, posters, visualization, formal and casual presentations

Weaker skills: R, Power BI, web-based user interfaces, Pachyderm, pySpark

Older skills: GIS, SAS and SAS modules, Tableau, Hive, Impala, Neo4J, MS Access, CSS, HTML

Certificates, Awards, Publications

Congressional Medal of Honor for Service in Antarctica

SAS Advanced Programming for SAS 9 certificate

Human Research Curriculum for Research Involving Data and Specimens

Human Participants Protection Education for Research Teams

EPA’s Scientific and Technological Achievement Award, Level III (received two)

Author or co-author on journal articles (18), presentations (40), reports, and manuals

Courses

Scalable Machine Learning (Spark), BerkleyX, EdX

Introduction to Big Data with Apache Spark, BerkleyX, EdX

Cluster Analysis in Data Mining, University of Illinois at Urbana-Champaign, Coursera

R Programming, John Hopkins, Coursera

Web Intelligence and Big Data, IIT, Coursera

An Introduction to Interactive Programming in Python, Rice University, Coursera

Topics in Artificial Intelligence, Jet Propulsion Laboratory, Pasadena CA

Hadoop and the Amazon Cloud, Big Data University

Other

Air quality sample collection, Ground Zero, NY, NY

Optical oceanographic measurements, Antarctica

Volunteer analyst with the SumAll Foundation

US Government Secret Clearance, 2010, inactive