Casson Stallings, Ph.D.

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

Career Summary

Data scientist with more than 20 years of proven experience in quantitative analysis and software development. Extensive and diverse experience wrangling and analyzing data. A successful history of consulting with internal and external clients in business, universities, and the government.

Highlights

* Technical lead on numerous projects; provided technical direction and facilitated communication and collaboration between clients, partners, and contractors.
* 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.
* Implemented a Neo4J graph database of people, companies, and financial organizations in the startup community, used Gephi and Python to investigate acquisitions among these start-ups.
* Developed methodology to increase the classification accuracy of imagery by combining geo-referenced vector data and maximum likelihood spectral classification.

Experience

Senior Data Scientist / Metabolon **2017 to Present**

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.

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 statistical 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.
* Supported data-driven sample site selection programs, carried out air quality sampling at Ground Zero.

Data Scientist, FutureAdvisor **2014 to 2015**

* Analyzed company’s database to provide executives with a better understanding of their customer demographics and trends in customer stock holdings and risk tolerance. Identified significant weaknesses in free customer diversification compared to managed portfolios which through customer education could be used to convert free to premium customers.

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 and 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
* Data Science Fellow, Zipfian Academy (now Galvanize), San Francisco CA

Skills

Techniques: Statistics, machine learning, data mining, modeling, exploratory and spatial analysis

Programming: SAS, SAS/MACRO, Python, R, Fortran

Analysis: SAS/STAT, SAS/GRAPH, Tableau, Spark, scikit learn, Seaborn, IPython, Zeppelin,

Spatial: ArcGIS, ArcView, Manifold

Databases: Hive, Impala, SQL, MySQL, Redshift/Postgres, Neo4J, Geodatabase, MongoDB, MS Access

UX: SAS AF/SCL, CSS, HTML, Avenue

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

Certificates, Awards, Publications

SAS Advanced Programming for SAS 9 certificate

Human Research Curriculum for Research Involving Data and Specimens

Human Participants Protection Education for Research Teams

Congressional Medal of Honor for Service in Antarctica

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

Author or co-author on many journal articles, presentations, and reports

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

Volunteer analyst with the SumAll Foundation

US Government Secret Clearance, 2010, inactive

Air quality sample collection, Ground Zero, NY, NY

Optical oceanographic measurements, Antarctica