

Christopher Galbraith

galbraic.github.io | [galbraic](https://github.com/galbraic) | [galbraic](https://www.linkedin.com/in/galbraic/)

Experience

Obsidian Security

MACHINE LEARNING RESEARCHER

Newport Beach, CA

Jul 2018 - Present

- Perform exploratory data analysis to identify and formulate problems where statistical models may offer appropriate solutions.
- Design and implement data models and pipelines to support new services used by models, security researchers, and front-end developers.
- Ensure and maintain quality of data transformations and models deployed in production from ingestion to what is shown in the final product.

DATA SCIENCE INTERN

Jun 2018 - Jul 2018

- Developed and deployed the company's first anomaly detection model.

Cylance Inc.

DATA SCIENCE INTERN

Irvine, CA

Jun 2017 - Sep 2017

- Developed and implemented a probabilistic model for real-time anomaly detection in network traffic data on endpoints.

University of California, Irvine

Irvine, CA

DATA SCIENCE INITIATIVE INSTRUCTOR & ORGANIZER

Oct 2016 - June 2018

- Designed and instructed short courses, *Introduction to R* and *Introduction to Data Analysis with R*, on the fundamentals of statistical modeling—from the basics of the R programming language to exploratory data analysis through visualizations to linear and logistic regression.
- Organized a Climate Science Hackathon focused on data visualization and machine learning.
- Created and maintained an SQL database of participant and instructor information on short courses for the Data Science Certificate program.

Education

University of California, Irvine

Irvine, CA

DOCTOR OF PHILOSOPHY, STATISTICS

May 2020

MASTER OF SCIENCE, STATISTICS

Jun 2016

South Dakota State University

Brookings, SD

BACHELOR OF SCIENCE, MATHEMATICS (MINORS IN COMPUTER SCIENCE & STATISTICS)

May 2014

Technical Skills

Python (Numpy, Pandas, Scikit-Learn, Pyspark), Scala, Spark, R (R Tidyverse), SQL (SQLite, PostgreSQL, MySQL), Databricks, Jupyter Notebook, Amazon Web Services (S3, EC2, RDS), GraphQL, ElasticSearch, Kubernetes, Docker, Git, Bash, JSON, XML, REST, SOAP

Projects

Anomaly Detection in Enterprise Cloud Environments

- Implemented and deployed an unsupervised model to detect anomalous activity sessions in enterprise cloud environments, including building featurization pipelines with Spark SQL and model training at scale with Spark ML pipelines.
- Monitored performance across customer environments to ensure proper model specification and make adjustments to feature pipelines and training methods over time.

Workday Service Connection

- Collaborated with subject matter experts to identify relevant security issues in customers' Workday tenants, then worked with backend engineers to design a system that collects the data necessary to solve said problems via REST and SOAP endpoints.
- Performed exploratory data analysis with Databricks and used the findings to create data transformation pipelines in Python and Spark that hydrates ElasticSearch indices which front-end developers used to display findings to customers.

Forensic Analysis of User-Generated Event Data

- Collaborated with forensic science practitioners to develop relevant statistical methods for assessing same-source forensic questions with applications to both temporal and spatial event data.
- Analyzed both simulated and real-world event data from a variety of sources (Twitter, web browsing, authentication, and mobile device data).
- Emphasized individual-level models for event data (mixture models with components for the population and individual).
- Published results in peer-reviewed journals with corresponding open source software (**assocr** and **spatial-assocr**).

Publications

- **C. Galbraith** (2020). Statistical methods for the forensic analysis of user-event data. PhD Thesis, UC Irvine.
- **C. Galbraith**, P. Smyth & H.S. Stern (2020). Statistical methods for the forensic analysis of geolocated event data. *Digital Investigation*, 33, 301009.
- **C. Galbraith**, P. Smyth & H.S. Stern (2020). Quantifying the association between discrete event time series with applications to digital forensics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183:3, p 1005–1027.
- **C. Galbraith** & P. Smyth (2017). Analyzing user-event data using score-based likelihood ratios with marked point processes. *Digital Investigation*, 22, p S106–S114.