SPENCER MATTHEWS

PhD Student & Data Scientist

spencer.matthews@uci.edu (916) 960-7531 https://srmatth.github.io

EXPERIENCE

CLA (CliftonLarsonAllen, LLP) • Graduate Data Science Intern • *October 2023 – Present*

- Designed and implemented processes for automated extraction of information from 5 different tax forms using large language (OpenAI) and computer vision (Azure Document Intelligence) models
- Advised agricultural clients and used results from statistical models built on their historical data to increase crop yields which led to over $250,000 of additional revenue per growing season
- Created a custom PDF "chunking" algorithm to aid in retrieval-augmented generation for an internal chat bot
- Used internal data to create a pricing tool so project billers better understand how to price their engagements

Bayer Crop Science • Data Science Intern • New Business Models • *June 2022 – September 2023*

- Cleaned, explored, and analyzed data from 4 large public datasets focused on crop insurance policies and claims, culminating in presenting relevant business insights and creating data dictionaries
- Performed *in silico* data experiments and presented the results to business partners to facilitate decision making
- Used economic theory to evaluate the desirability of new business models for the farmer and Bayer

CLA (CliftonLarsonAllen, LLP) • Data Science Intern • *January 2020 – September 2021*

- Developed R Shiny applications within the Golem framework for 5 internal projects which experienced heavy production use by up to 300 concurrent users
- Contributed to 4 internal R packages whose goal was seamless operations with Azure tools
- Automated PowerPoint creation with R to populate template slide decks with matrix-specific information
- Extracted and cleaned numerous datasets from firm databases so business leaders could assess performance

Brigham Young University • Research Assistant • Dr. Brian Hartman • *December 2019 – September 2021*

- Extended the SHAP model explainability algorithm to two-part models so machine learning methods can be used for prediction in regulated industries; this included publishing an R package on the CRAN
- Authored two papers explaining our new methodology and presented the results at the 24th IME conference
- Created, organized, and maintained two open-access public code repositories

EDUCATION

University of California – Irvine • *Irvine, California*
Ph.D. in Statistics (expected June 2026)
M.S. in Statistics (June 2023)
2022 Newcomb Award Winner

Brigham Young University • *Provo, Utah*
B.S. in Statistics (December 2020)
Summa Cum Laude

Notable Courses: Advanced Survival Analysis, Bayesian Statistics, Spatial Statistics, Advanced Probability Series, Financial Mathematics, Economics, Data Structures and Algorithms, Introduction to SQL, Introduction to Unix

Other Certifications: SOA Exam P (July 2024), SOA Exam FM (August 2024), Portfolio Construction and Analysis with Python (Coursera, July 2023), Data Visualization Professional Certification (LinkedIn Learning, June 2021)

SKILLS

Programming Languages: Python, R (advanced); HTML, CSS (intermediate); JavaScript, SQL, C++, Unix (proficient)

Frameworks/Software: Git/Github, Microsoft Azure, Posit Workbench, Databricks/Spark, AWS S3, Camunda, Docker

Languages: English (native), Spanish (professional fluency)