

A versatile data scientist skilled in implementing and fine-tuning advanced models, with a strong foundation in probability theory and applied statistics. Proficient in developing machine learning models and creating algorithms to manage large datasets while ensuring optimal performance. Adept at simplifying technical concepts for diverse audiences using straightforward explanations, visual tools, and relatable examples.

Skills

Programming & Tools

Python | R | MATLAB | JavaScript | HTML | CSS | PostgreSQL | MySQL | SQLite | RSQLite | Jupyter | BeautifulSoup | Numpy | Pandas | Scikit-learn | TensorFlow | PyTorch | Keras | XGBoost | LightGBM | sqlalchemy | ArcGIS | Stata | Git | LaTeX | Markdown | Retrieval-Augmented Generation (RAG) | Local Large Language Models

Data Processing & Visualization

Data Pre-processing | Feature Engineering | Hyper-parameter Tuning | Cross-validation | Sentiment Analysis | Topic Modeling | Text Classification | Data Cleaning | Data Normalization | Missing Data Imputation | Data Aggregation | Advanced Integration Techniques | NLTK | SpaCy | ggplot2 | plotly | Advanced NLP Techniques

Statistical & Machine Learning Techniques

Bayesian Model Averaging | Random Forest | Support Vector Machine | Gradient Boosting Machines | k-Nearest Neighbors | Linear Regression | Logistic Regression | Maximum Likelihood Estimation | Causal Inference | Propensity Score Matching | Regression Discontinuity Design | Difference-in-Differences | Reinforcement Learning from Human Feedback | Conjoint Experiment Design | Panel Data Analysis | Robust Regression | Bootstrapping | Genetic Matching | Experimental and Observational Data Integration

Experience

Data Science Specialist

April 2023 - present
Remote

Scale AI

- Contributed to a reward model of reinforcement learning from human feedback (RLHF) to align with client-specific needs
- Emphasized improvements in accuracy, conversational fluidity, and user engagement while maintaining strict writing style and structure protocols across various domains
- Incorporated sophisticated AI capabilities, including semantic comprehension and topic adherence to optimize model performance
- Formulated detailed training scripts for a range of data science applications, strengthening the chatbot's proficiency in code interpretation and generation

Computational Scientist

August 2017 — May 2024

Washington University in Saint Louis

Saint Louis, MO

- Implemented advanced ensemble machine learning techniques, such as Bayesian model averaging and random forests
- Developed an R package (EBMAforecast) for predicting heterogeneous treatment effects

- Conducted extensive data pre-processing, exploratory analysis, handling of missing values, outlier detection, and feature engineering
- Optimized model performance through rigorous hyper-parameter tuning, cross-validation, and appropriate feature set selection

Data Science Instructor

Washington University in Saint Louis

June 2020 — June 2022

Saint Louis, MO

- Developed curriculum teaching Python programming, data analysis techniques, Git, and JupyterLab
- Taught Python fundamentals, data manipulation, and visualization using `pandas`, `numpy`, `seaborn` and `matplotlib` libraries
- Supervised hands-on projects, highlighting the use of GitHub for version control and JupyterLab for interactive analysis
- Adapted teaching methods for diverse learners, preparing students to confidently employ data-driven approaches

Education

Ph.D. in Political science	Washington University in Saint Louis	2024
M.A. in Political Science	Saint Louis University	2017
B.A. Political Science (Cum Laude)	Saint Louis University	2016

Publications

Guess, Andrew, Dominique Lockett , Benjamin Lyons, Brendan Nyhan, Jacob M. Montgomery, and Jason Reifler. 2020. “Fake news’ may have limited effects beyond increasing beliefs in false claims.” The Misinformation Review. [Link to article.](#)

Edelson, Laura, Dominique Lockett , Jacob Montgomery, Damon Mccoy, Tobias Lauinger, Celia Guillard“US Public Opinion Towards Platform Regulation of Political Advertisements: Discontent and Consensus for Reform” (Forthcoming)

Lockett, Dominique. Cognitive Landscapes: Argument Evaluations, Misinformation Corrections, and Racial Attitudes in Modern Media. [Link to book.](#)

References

Jacob Montgomery, PhD	Chair of Division of Computational and Data Sciences, WashU	Contact info.
Betsy Sinclair, PhD	Chair of Political Science, WashU	Contact info.

Projects

Hosting Local Large Language Model

[Deep Learning, Data Management, Local Hosting, User Interface]

Present

Saint Louis, MO

- Applied `PyTorch` and `NumPy` for the implementation and training of a large language model, leveraging pre-trained weights and GPU acceleration for efficient computation and performance
- Implemented Retrieval-Augmented Generation (RAG) to enhance the model’s capability in retrieving and generating contextually relevant information from extensive external knowledge bases
- Engineered efficient indexing and search algorithms to significantly improve retrieval accuracy and speed within the RAG framework

- Optimized storage and retrieval processes to handle large-scale data efficiently, ensuring seamless data management and accessibility

News Discourse Analysis via Advanced NLP Techniques

[Big Data, Machine Learning, Data Engineering, Political Communication]

June 2023

Saint Louis, MO

- Scraped and assembled a SQL database containing billions of tweets and replies from 30 news outlets to examine communication patterns and styles
- Designed an automated ETL pipeline using Python packages such as `nltk` and `numpy`
- Utilized advanced NLP techniques such as sentiment analysis, topic modeling, and text classification using packages like `BERT`, `gensim`, and `scikit-learn`

Forecasting the 2020 Presidential Election

[Machine Learning, Bayesian Inference, Time-Series Analysis, Forecasting]

July 2019

Saint Louis, MO

- Conducted extensive data pre-processing, feature engineering, and model evaluation using cross-validation techniques
- Developed a Bayesian model averaging approach to combine predictions from multiple forecasting models
- Applied the ensemble method to the 2020 presidential election, resulting in highly accurate forecasts
- Designed and implemented an MCMC sampling scheme for rigorous statistical inference and uncertainty quantification

Political Advertisement Perception Study with Advanced Analytics

[Data Engineering, Data Visualization, Conjoint Experiment, Political Advertising]

December 2021

Saint Louis, MO

- Deployed a conjoint experiment to understand user preferences and perceptions of political ads
- Leveraged expertise in non-parametric analysis with tools such as `cjoint`, `coefplot`, and `gmodels` packages for advanced statistical modeling and analysis
- Ensured accurate interpretation of results using statistical concepts such as weighted means, confidence intervals, and hypothesis testing
- Designed and developed informative visualizations with Python packages such as `ggplot2` and `plotly` packages and `ggplot2` in R

Analyzing Misinformation Effects in Political Participation

[Research Design, Data Analysis, Machine Learning, Data Visualization, Misinformation]

July 2022

Saint Louis, MO

- Conducted a comprehensive multi-university survey to explore the impacts of fake news on political participation and identify effective strategies to prevent misperceptions
- Developed and implemented advanced statistical analyses using R, including regression models and group comparisons, to assess the relationship between fake news exposure and political behavior
- Created sophisticated data visualizations using `ggplot2`, including histograms, density plots, bar charts, and time series plots, to illustrate key findings on fake news exposure and its effects
- Performed focused analyses on the impact of fake news exposure on political polarization and the effectiveness of fact-checking interventions, presenting results through custom-designed plots
- Ensured reproducibility and transparency by providing detailed R Markdown scripts for data analysis and visualization in the project's GitHub repository

- Devised a novel survey experiment to assess the efficacy of misinformation corrections using advanced statistical techniques
- Implemented a pilot survey and deployed the final survey to a diverse, nationally-representative sample
- Analyzed survey data, employing R packages such as `dplyr`, `tidyverse`, and `ggplot2` for data manipulation, visualization, and application of multivariate statistics

Awards and certificates

Research seed grant	Washington University in Saint Louis	2020
Diversity fellowship	Saint Louis University	2016
Fundamentals of G.I.S.	University of California, Davis	2020

Relevant Coursework

Quantitative Political Methodology II 2020

- Advanced course focused on sophisticated statistical analysis methods for computational scientists
- Emphasized maximum likelihood estimation for various scenarios, including cross-sectional, time series, and non-parametric bootstrapping
- Materials: *All of Statistics: A Concise Course in Statistical Inference*, Larry Wasserman; *R Programming for Data Science*, Roger D Peng; *Bayesian Data Analysis (Third Edition)*, Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin.; Taught by Jacob Montgomery

Computational Social Science 2020

- Explored various data types in social science, including networks, text, audio, images, and videos
- Focused on both mechanistic and probabilistic approaches to supervised and unsupervised learning
- Materials: *Pattern Recognition and Machine Learning*, Christopher Bishop; *A Course in Machine Learning*, Hal Daumé; *The Elements of Statistical Learning*, Jerome Friedman, Trevor Hastie, Robert Tibshirani; Taught by Christopher Lucas

Maximum Likelihood Estimation 2019

- In-depth focus on MLE principles, including probability theory, likelihood functions, and properties of estimators like consistency and efficiency
- Comprehensive study of generalized linear models using MLE, covering exponential family distributions, link functions, logistic and Poisson regression
- Advanced MLE topics: handling categorical data, overdispersion in count data, model selection criteria (AIC, BIC), model fit assessment and diagnostics
- Materials: *Generalized Linear Models*, Peter K Dunn, Gordon K Smyth; Taught by Christopher Lucas

Causal Inference

2019

- Deep exploration of causal inference theories, focusing on counterfactual reasoning, potential outcomes, and causal diagrams
- Study of experimental design principles, including randomized trials, natural and field experiments
- Exploration of observational techniques: propensity score matching, regression discontinuity, difference-in-differences, instrumental variables
- Advanced statistical methods for causal estimation: structural equation modeling, mediation analysis, sensitivity analysis; Taught by Julia Park

Applied Statistical Programming

2018

- Introduced object-oriented programming, functional programming paradigms, and efficient data manipulation
- Covered topics such as debugging, profiling, as well as package development and contribution to open-source projects
- Emphasized statistical meta-skills like data cleaning, transformation, visualization, and implementation of various statistical models and algorithms
- Materials: *R for Dummies*, de Vries and Meys; *Advanced R*, Hadley Wickham; Taught by Jacob Montgomery

Theories of Individual and Collective Choice I (Game Theory)

2018

- Study of rational choice theory, delving into strategic decision-making processes, utility maximization, and behavioral strategy
- Game-theoretic models: extensive and normal form games, Nash equilibrium concepts, repeated and dynamic games
- Analysis of cooperative game theory, focusing on coalition formation, bargaining theories, and the Shapley value
- Advanced topics: evolutionary game theory, Bayesian games, and information asymmetry in strategic interactions
- Materials: *Game Theory: An Introduction*, Steven Tadelis; Taught by Keith Schnakenberg

Quantitative Political Methodology I

2017

- Explored mathematical underpinnings of linear regression models, exploring both scalar and matrix representations
- Covered extensive topics including estimation techniques, inference methods, assumptions of linear models, diagnostic procedures, and the implementation of these concepts in statistical computation
- Special focus on understanding the Gauss-Markov theorem, least squares estimation, multicollinearity, heteroskedasticity, and model specification errors
- Materials: *Linear Models with R*, Julian J. J. Faraway; Taught by Guillermo Rosas

Mathematical Modeling

2017

- Explore advanced mathematical concepts, particularly matrix algebra and calculus, within the framework of economic modeling
- Topics include matrix operations, determinants, eigenvalues and eigenvectors, and their applications in solving linear systems
- Covered single-variable and multivariate calculus, including a detailed study of limits, continuity, differentiation, and integration
- Materials: *Mathematics for Economists*, Pemberton and Rau; Taught by Randy Calvert

Research Design

2017

- Explored the application of the philosophy of science in the social sciences
- Topics included research methodologies, hypothesis formation and testing, the structure of scientific inquiry, and the principles of logical reasoning
- Addressed the challenges of causality, including the design of experiments and observational studies, and the use of statistical methods for causal inference
- Materials: *Political Science and the Logic of Representations*, Kevin A Clarke and David M Primo; *The Logic of Real Arguments*, Alec Fisher; Taught by Matt Gabel