github
website
linkedin

Dominique Lockett

Data Scientist/Ph.D Candidate

St. Louis, MO
6189546277
dee_lockett@hotmail.com

A versatile data scientist with a comprehensive skill set, combining expertise in the implementation and fine-tuning of state-of-the-art models like GPT with an adept understanding of statistics and probability theory. Experienced in the development of machine learning techniques and proficient in handling vast data structures and optimizing their performance. Skilled in simplifying and conveying technical concepts to varied audiences via straightforward explanations, visual tools, and relatable examples. Poised to thrive in settings seeking a data-centric professional with a deep grounding in statistical theories, machine learning, deep learning methodologies, and a track record of applying them in diverse domains, from academia to AI-driven business applications.

Skills

Programming Languages	Python | R | MATLAB
Web Technologies & Databases	JavaScript | HTML | CSS | PostgreSQL | MySQL | SQLite | RSQLite
Frameworks & Libraries	Jupyter | BeautifulSoup | Numpy | Pandas | Scikit-learn | TensorFlow | PyTorch | Keras | XGBoost | LightGBM | sqlalchemy
Tools, Software & Formatting	ArcGIS | Stata | Git | LaTeX | MarkDown

Experience

Data Science Expert — **April 2023 - present**
Scale AI — Saint Louis, MO

- Contributed to a reward model of reinforcement learning from human feedback (RLHF) to align with client-specific needs, emphasizing improvements in accuracy, conversational fluidity, and user engagement
- Oversaw quality assurance and provided critical evaluation as a review specialist for RLHF projects, ensuring adherence to the highest standards
- 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
- Effectively synthesized information while maintaining strict writing style and structure protocols across various domains

Python Instructor — **June 2020 — June 2022**
Washington University in Saint Louis — Saint Louis, MO

- Developed a curriculum teaching Python programming, data analysis techniques, Git, and JupyterLab for public health master's students
- 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 address public health problems with data-driven approaches

Research Assistant **June 2018 — December 2022**
Washington University in Saint Louis Saint Louis, MO

- Implemented advanced ensemble machine learning techniques, such as Bayesian model averaging and random forests, to develop an R package (`EBMAforecast`) for predicting heterogeneous treatment effects in political research
- Conducted extensive data pre-processing, exploratory analysis, handling of missing values, outlier detection, and feature engineering to ensure high-quality data input for model development
- Optimized model performance through rigorous hyper-parameter tuning, cross-validation, and appropriate feature set selection, leading to improved predictive accuracy and robustness across contexts
- Collaborated with interdisciplinary teams of political scientists, statisticians, and computer scientists to apply methodologies and communicate complex findings effectively

Projects

Hosting Local Large Language Model **July 2023-October 2023**
Deep Learning, Data Management, Local Hosting, User Interface

- Utilized `PyTorch` and `NumPy` for the implementation and training of the large language model, leveraging pre-trained weights and GPU acceleration for efficient computation
- Became proficient in the theories related to the general structure of LLMs, which includes neural network components like recurrent, feedforward, embedding, and attention layers
- Hosted the project on a public repository and provided basic guidance to allow other users to implement LLM
- Designed a local hosting environment that enables the deployment of the large language model within an on-site infrastructure
- Ensured efficient storage and retrieval processes, handling large-scale data through optimization techniques

News Discourse Analysis via Advanced NLP Techniques **June 2020-June 2023**
Big Data, Machine Learning, Data Engineering, Political Communication

- Constructed a large-scale dataset of Twitter data from 30 news outlets to examine communication patterns and styles, involving billions of tweets
- Designed and implemented a scalable ETL pipeline using Python packages such as `nltk` and `numpy` to develop efficient pre-processing in the management of big data
- Applied generalized linear mixed models and Bayesian statistics to analyze communication patterns, utilizing advanced NLP techniques such as sentiment analysis, topic modeling, and text classification using packages like `BERT`, `gensim`, and `scikit-learn`
- Demonstrated proficiency in handling API requests, pagination, and error management with `Tweepy`, ensuring a robust data collection process from Twitter
- Efficiently utilized `SQLAlchemy` ORM capabilities with `SQLite` for database storage and retrieval, employing features like declarative mapping, query composition, and connection pooling for optimal performance
- Presented complex quantitative analyses and findings to key stakeholders in a clear, concise, and actionable manner

Forecasting the 2020 Presidential Election **July 2019**

Machine Learning, Ensemble Methods, Bayesian Inference, Time Series Analysis, Election Forecasting

- Conducted extensive data pre-processing, feature engineering, and model evaluation using cross-validation techniques to optimize model performance and ensure the validity of the forecasts
- Developed a Bayesian model averaging approach to combine predictions from multiple forecasting models, leveraging expertise in experimental design and multivariate statistics
- Applied the ensemble method to the 2020 presidential election, resulting in a highly accurate and precise forecast with a 95% credible interval using 11 historical out-of-sample forecasts
- Designed and implemented an MCMC sampling scheme in `R` for rigorous statistical inference and uncertainty quantification, enabling a more comprehensive understanding of model performance and potential biases

Political Advertisement Perception Study with Advanced Analytics **December 2021**

Data Engineering, Data Visualization, Conjoint Experiment, Political Advertising

- Conducted a conjoint experiment to understand user preferences and perceptions of political ads, leveraging expertise in non-parametric analysis with tools such as `cjoint`, `coefplot`, and `gmodels` packages for advanced statistical modeling and analysis
- Examined political and social issues survey data with `descr`, `labelled`, and `haven` packages to identify key trends and insights in public opinion
- Ensured accurate interpretation of results using statistical concepts such as weighted means, confidence intervals, and hypothesis testing
- Designed and developed informative visualizations with `ggplot2` and `plotly` packages, effectively communicating complex findings
- Utilized strong communication competencies to present actionable insights to extended teams and small groups of key stakeholders

Analyzing Misinformation Effects **July 2022**

Research Design, Data Analysis, Machine Learning, Data Visualization, Publication, Misinformation

- Developed a novel research framework integrating experimental and observational data to investigate misinformation effects on public opinion and behavior
- Employed machine learning techniques such as regression and generalized linear mixed models, classification, and clustering algorithms to evaluate causal factors and predict misinformation susceptibility
- Ensured validity and generalizability by using nationally representative samples and cross-validation techniques in the R
- Fostered transparency and reproducibility by submitting research data, code, and documentation to the Harvard Dataverse repository, enabling further collaboration and research in the field

Measuring Effectiveness of Misinformation Corrections: Design and Implementation **September 2022**

Research Design, Data Analysis, Data Visualization, Grant Writing, Misinformation

- Devised a novel survey experiment to assess the efficacy of misinformation corrections using advanced statistical techniques, including non-linear mixed models and Bayesian statistics
- Implemented a pilot survey and deployed the final survey to a diverse, nationally-representative sample, ensuring generalizability of the results

- Analyzed survey data using R programming, employing packages such as `dplyr`, `tidyverse`, and `ggplot2` for data manipulation, visualization, and application of multivariate statistics
- Developed data-driven insights on the effectiveness of various misinformation correction techniques and presented findings through clear and informative visualizations
- Contributed to grant writing efforts, securing funding for future research on misinformation and its impact on public opinion and behavior
- Effectively communicated complex quantitative analyses to extended teams and key stakeholders, utilizing strong communication competencies to present findings in a clear, concise, and actionable manner

Education

Ph.D. in Political science	Washington University in Saint Louis	August 2017 — July 2024
M.A. in Political Science	Saint Louis University	August 2016 — May 2017
B.A. Political Science (Cum Laude)	Saint Louis University	August 2013 — May 2016
A.A. Communication	Ivy Tech Community College	August 2009 — May 2012

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.

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. Using Objectivity to Improve Argument Evaluations. (Forthcoming)

Awards and certificates

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

Coursework

Quantitative Political Methodology II **2020**

- Advanced political methodology course focusing on sophisticated statistical analysis methods for political and social scientists.
- Featured essential materials from "Linear Models with R" by Julian Faraway.
- Explored topics included maximum likelihood estimation across cross-sectional, time series, and non-parametric bootstrapping scenarios.
- Taught by Jacob Montgomery (advisor).

Computational Social Science **2020**

- Explored diverse data types such as networks, text, audio, images, and videos.
- Used primary resources from "Pattern Recognition and Machine Learning" by Christopher Bishop, "A Course in Machine Learning" by Hal Daumé, and "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" by Jerome Friedman, Trevor Hastie, Robert Tibshirani.
- Started with mechanistic approaches to supervised and unsupervised learning, progressing to statistical inference with probabilistic interpretations.
- Taught by Christopher Lucas.

Maximum Likelihood Estimation **2019**

- Focused on generalized linear model estimation.
- Provided practical exposure to link functions for various models, including multinomial and unordered models, ordered outcome models, duration models, and count models.
- Taught by Julia Park.

Causal Inference **2019**

- Introduced theoretical frameworks for causality and the empirical tools used in estimating causal effects.
- Participants learned and applied skills related to outcomes, causal graphs, experiments, matching, regression, difference-in-differences, instrumental variables, sensitivity analysis, and regression discontinuity.
- Taught by Julia Park.

Applied Statistical Programming **2018**

- Developed programming skills in R.
- Provided exposure to the fundamentals of object-oriented programming.
- Used primary materials from "R for Dummies" by de Vries and Meys and "Advanced R" by Hadley Wickham.
- Focused on teaching foundational meta-skills from computer science and statistics.
- Taught by Jacob Montgomery (advisor).

Theories of Individual and Collective Choice I (Game Theory) **2018**

- Introduced rational choice theory.
- Covered spatial theory of electoral competition, cooperative game theory, and general equilibrium theory.
- Used "Game Theory: An Introduction" by Steven Tadelis as its primary material.
- Taught by Keith Schnakenberg.

Quantitative Political Methodology I **2017**

- Utilized "Linear Models with R" by Julian Faraway.
- Explored the fundamentals of linear regression models in both scalar and matrix forms.
- Problem sets focused on estimation, inference, specification, diagnostic tools, data management, and statistical computation.
- Taught by Guillermo Rosas.

Mathematical Modeling **2017**

- Led by Randy Calvert.
- Featured "Mathematics for Economists" by Pemberton and Rau.
- Covered single-variable calculus and portions of multi-variate calculus, linear algebra, and probability theory.
- Topics included sets and relations, probability, differential calculus and optimization, difference equations, and linear algebra.

Research Design **2017**

- Explored the philosophy of science and its applications in political science research.
- Focused on scientific knowledge, principles, and standards for evaluation and argumentation in social scientific research.
- Used primary materials such as "Political Science and the Logic of Representations" by Kevin A. Clarke and David M. Primo, and "The Logic of Real Arguments" by Alec Fisher.
- Taught by Matt Gabel.