Andrew Fogarty
seekinginference.com
apfogarty@berkeley.edu
434-806-2891

Education

University of California, Berkeley
Master of Information and Data Science — *Expected: April 2021*

University of Virginia
Master of Political Science — *2015-2017*

Methods

Statistics: Linear, Unordered Categorical, Ordered Categorical, Binary, Count, Survival, Time Series, Time Series Cross-Section, Multiple Imputation, Numerical Optimization

Machine Learning: Nearest Neighbors, Decision Trees, Support Vector Machines, Dimensionality Reduction, Clusters

Deep Learning: MLP, Regularization, Normalization & Initialization, Optimization, Distributed

Natural Language Processing: CNN, Capsules, T5, Entity Embeddings, BERT, Generation, Summarization, Semantic Search

Qualitative: Research Design and Causal Inference, Comparative Methods, Essential Empirical Methods

Technologies: Python, R, AWS, Spark, Petastorm, Horovod, DataBricks, SQL, PyTorch, AWS-SageMaker

Key Projects

PetaFlights: Placed model into production with TorchServe in an academic setting using a 31 million row deep learning time series tabular data set, processing 450k rows/minute on CPU across 8 workers using state-of-the-art Petastorm, Horovod, PyTorch, Spark, and DataBricks.
Taliban Mobility: Conducted causal inference research on the determinants of a military group's mobility across geographic units and time using a time series cross-sectional linear model, treatment and control case studies, and custom maps. This research influenced the President's Daily Brief, was cited in defense of the U.S. National Intelligence Estimate's key judgments, and was selected for presentation at a data science conference.
Cities Under Attack: Proposed causal explanations for why cities in Afghanistan are attacked using a time series logit model, treatment and control case studies, and dashboards to animate time series data. This research was selected for presentation at a data science conference.
Enemy at the Gates: Delivered novel cross-sectional research through interactive graphics and visual executive summaries that estimated the most important factors associated with the deaths of tens of thousands of allied foreign military forces in Afghanistan.
Natural Language Propaganda: Created novel NLP data set from scratch on 11,000 Taliban propaganda messages from 2014-2020 for use in classification, text generation, and summarization tasks on an array of transformers and baseline models including: BERT, T5, and GPT-2 in PyTorch.

Work Experience

National Ground Intelligence Center — Charlottesville, VA
Data Scientist — *2010 - Present*

Technical: Provided technical consultation to my colleagues on methodological matters such as overcoming non-random assignment to treatment, data summation, data visualization, data gathering strategies, hypothesis testing, and research design.

Customer Service: Worked closely with government and military leaders at tactical, senior, and executive levels to translate customer interests into actionable research on dozens of occasions.

Leadership: Led individuals and groups of analysts at all stages of analysis from research design to data collection, data coding, and drafting in order to produce a series of insightful and superior collaborative assessments.

Invited Data Science Conference Presentations

5th Annual Data Science Technical Exchange — Dulles, VA
Central Intelligence Agency — *September 2019*

ConMap Meet: Conflict Mapping — Springfield, VA
National Geospatial-Intelligence Agency — *May 2019*

4th Annual Data Science Technical Exchange — Bethesda, MD
Central Intelligence Agency — *September 2018*