Zachary M. Jones, Ph.D.

zmjone2992@gmail.com
(912) 401-8711
LinkedIn

Summary

Data scientist with a background in social science, machine learning, statistics and software development, with experience identifying org-wide problems and building/leading cross-team collaborations to solve them.

Experience

- Senior Applied/Data Scientist, *Amazon* (10/23 –)

    - Developed automatic appeal evaluation and model explanation system based on technical signals, LLM summaries, Shapley values, and fraud tagging pipeline
    - Developed new fraud tagging pipeline integrated with payment instrument fraud detection and login integrity teams, doubling labelled fraud and resulting in upstream enforcements
    - Development of/standards for AI-first ML development environment for team (orchestrating end-to-end development)
    - Developed session-level risk model resulting in 2x impact versus previous best, saving $19.2M/yr in fraud leakage (+$9.6M vs. status-quo)
    - Developed new enforcement decisioning heuristic and publisher risk model improving $ weighted precision by 5.5x, and recall by 2x, reducing fraud royalty payments by $1.57M/year and false positive impact by $10.3M
    - Led deprecation of legacy blocklisting system reducing false positive impact by $34M/year
    - Developed automation of monthly retroactive fraud enforcement process reducing manual effort by 1 day/month

- Senior Research Scientist, *Meta* (12/21 – 6/23)

    - Led multi-team technical product development for automated fake advertiser checks, developed measurements of adversarial adaptation to advertiser account disables and leveraged them to improve system efficacy with $10M+/year business impact
    - Developed, experimentally tested and launched multiple iterations of dynamic decisioning infrastructure which makes cost and effectiveness aware malicious advertiser restriction decisions with $100M+/year impact

- Senior Analyst, *Institute for Health Metrics and Evaluation* (5/19 – 12/21)

    - Developed an end-to-end forecasting system (a Python library) that was a major performance improvement over the status-quo, producing more accurate forecasts with uncertainty bounds
    - Built a data pricing application which was used in business development to rapidly iterate on sales packages
    - Estimated and forecasted lung cancer incidence/prevalence/mortality by histological subtype using incomplete and inconsistent data, which I reconciled and used for a publication

- Moore/Sloan Data Science Postdoctoral Fellow, *eScience Institute* (8/17 – 5/19)

- – Developed machine learning interpretation software, published in the *Journal of Machine Learning Research*, the *Journal of Open Source Software*, the *R Journal*, along with applications published in the *Journal of Politics*

- Google Summer of Code Fellow, `Machine Learning in R` (5/15 − 8/15)

  - – Implemented interpretability module in `mlr` based on dissertation research, which allows computation of feature importance in arbitrary dimensions agnostic of model type, used for model development, validation, and inference

- Graduate Research/Teaching, *Pennsylvania State University* (8/14 − 5/17)

  - – Developed a Monte-Carlo simulation of a network with arbitrary dependence between a latent network confounder and an observable feature, and showed that a class of latent network models in common use were unable to correctly adjust for the bias from the latent network confounder

- Graduate Research/Teaching, *University of Georgia* (8/11 − 5/14)

  - – Developed an ML model which ranked explanations for variation in cross-national violence based on their predictive value, resulting in a publication in the flagship journal
  - – Developed a Python library which scraped and cleaned all UN treaties for research

Education

- Ph.D. Political Science (statistical and computational methods), *Pennsylvania State University* (2017).

- M.A. Political Science, *University of Georgia* (2013).

- B.A. Political Science and Philosophy, *Georgia Southern University* (2010).

Selected Publications

1. mmpf: Monte-Carlo Methods for Prediction Functions, *The R Journal* (2018).

2. Is There More Violence in the Middle?, *American Journal of Political Science* (2018).

3. edarf: Exploratory Data Analysis using Random Forests, *Journal of Open Source Software* (2016).

4. mlr: Machine Learning in R, *Journal of Machine Learning Research* (2016).

5. Enhancing Validity in Observational Settings When Replication is Not Possible, *Political Science Research and Methods* (2016).

6. An Empirical Evaluation of Explanations for State Repression, *American Political Science Review* (2014).