Zachary M. Jones, Ph.D.

Remote // zmjone2992@gmail.com // zmjones.com // github.com/zmjones // (912) 401-8711

Summary

- data scientist with a strong background in social science, machine learning, statistics and software development
- experience identifying org-wide problems and building/leading cross-team collaborations to solve them
- production software development experience for machine learning/data pipelines with Python and R for research and industry applications
- experience presenting technical and non-technical material to academics and industry professionals

Experience

- Senior Data Scientist, *Amazon* (10/23 –)
- Senior Research Scientist, *Meta* (12/21 – 6/23)
  - Lead 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. Responsible for experimental design, analysis, product design, measurement, model development
  - Improved experimental methods and measurement quality across business integrity by designing numerous experiments and building new measurement pipelines to measure ML decision performance, as well as developing product launch criteria
  - 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
  - Lead multi-team effort to improve human review recall after identifying adversarial exploitation of the advertiser post-enforcement appeal system
- 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 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, smoothed, forecasted, and wrote a paper for publication based on
- Moore/Sloan Data Science Postdoctoral Fellow, *eScience Institute* (8/17 – 5/19)
  - Developed and published general purpose machine learning interpretation software, published in the Journal of Machine Learning Research, the Journal of Open Source Software, the R Journal

  - Applied interpretable ML software, and built an ML model to predict cross-national violence and its relationship with structural government features resulting in publication in a top-tier journal

- Google Summer of Code Fellow (5/15 – 8/15)
  - Implemented general purpose interpretability module based on dissertation research, which allows computation of feature importance in arbitrary dimensions agnostic of model type, used for model development, validation, and inference
  - General purpose work as an `mlr` developer; fixing bugs and improving stability
  - Implemented uncertainty estimation module using bootstrap/jackknife
  - Collaborated on the development of generic system for BayesOpt hyperparameter tuning

- 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)