Keith J. Yoder

Chicago, IL | (574) 215-9678 | kjyoder@gmail.com | linkedin:keithyoder | github:kjyoder

Summary: I am a data scientist with 13+ years of experience applying statistical modeling and machine learning complex problems in cognitive neuroscience and behavioral economics. I am dedicated to cross-functional collaboration aimed at leveraging technology to develop cutting-edge models that facilitate clinical care. I have used advanced analytic techniques in 13 first-author peer-reviewed publications in leading scientific journals, including three novel applications of probabilistic modeling.

DATA SCIENCE EXPERIENCE

SPARK Neuro Inc., New York, NY

Science Manager October 2023-Present

- Formalized scientific collateral procedures, reducing time to complete by 25%
- Supervised transition from 22 separate repositories into a monorepo, reducing release time by 85%
- Implemented Agile work estimation for issues, improved completion rate of commitments by 45%.

Senior Data Scientist July 2022-September 2023

- Built ETL pipelines to onboard and standardize 282 external EEG datasets, including visual inspection for data quality
- Integrated data from multiple amplifiers, improving model accuracy for detecting dementia by 5%
- Demonstrated utility of fractal distributions in distinguishing causes of dementia

Department of Psychology, University of Chicago Chicago, IL

Senior Research Analyst June 2021-June 2022

- Used independent component analysis to distinguish blink, ECG, and EMG from cortical signals in EEG recordings, trained and supervised 10 team members to implement ICA-based data cleaning
- Built SVM model to predict video content from EEG data collected during naturalistic viewing and identified 5% poorer prediction accuracy among women compared to men
- Reduced feature extraction and model training times for EEG studies by 70% by replacing Morlet wavelets with fast Fourier transforms to calculate spectral power density

Post-Doctoral Scholar July 2017-May 2021

- Developed and maintained in-house R/Python software for conducting multi-level analysis and with EEG and fMRI data, trained 12 team members to utilize it, created reproducible pipelines
- Integrated bash, Python, and MATLAB into ETL pipeline to download and preprocess MRI and behavioral data and facilitate QA checks, increasing efficiency by 80%
- Utilized distributed computing to decrease image processing time by 90%, from 144 days to 5 days

EDUCATION

University of Chicago Chicago, IL

Ph.D., Psychology (Integrative Neuroscience) 2017

M.A., Social Science 2011

Cornell University Ithaca, NY

B.S., Human Development 2009

TECHNICAL SKILLS

- Programming: Python (scikit-learn, matplotlib), R (tidyverse, RStan), MATLAB, bash
- Text and Web: SQL, Javascript, HTML, CSS, Markdown, LaTeX, Microsoft Office
- Data Analysis: Multi-level modeling, Bayesian statistics, machine learning, bootstrapping