



Matthew Epland

 matthew-epland •  mepland

Experience

Overjet

Senior Data Scientist

New York, NY

March 2023 – November 2023

- Developed a dimensional data warehouse and Looker dashboards to monitor new product launches
 - Used DBT, Airflow, and SQLFluff to create a reliable production environment

Komodo Health

Data Scientist

New York, NY

June 2019 – March 2023

- Planned and executed modeling projects for top 25 pharmaceutical companies
 - Projects included building rare disease classification models in Python and survival analysis models in R
- Assessed the value of potential new data sources
 - Analyzed medical and pharmacy claims data with PySpark running on AWS and Kubernetes
- Wrote efficient SQL queries against multi-TB tables on Snowflake, lead SQL optimization training sessions
- Contributed to the data science team's continuous improvement through quarterly seminar talks and mentorship

ATLAS Collaboration, Duke University

Research Assistant

Durham, NC

August 2014 – May 2019

- Hunted for new particles with the multi- b SUSY search team in $\mathcal{O}(\text{TB})$ of data, Ph.D. dissertation research
 - Improved search's sensitive area by $\sim 30\%$ through the application of a BDT classifier in XGBoost and Python

Data Science Educator

Teaching Assistant

2017 – 2019

- Data Science Pipeline and Critical Thinking Teaching Fellow in the 2019 Harvard Business Analytics program
- Data Visualization Teaching Assistant in the 2017 Fuqua Master of Quantitative Management (MQM) program

Education

Duke University

Ph.D. Physics

Durham, NC

March 2019

University of Minnesota, Twin Cities

B.S. Physics Summa Cum Laude, B.S. Mathematics

Minneapolis, MN

May 2014

Professional Activities

Open Source Projects

Contributor

September 2022 – Present

- Regular contributor to the `imodels` and `dtreeviz` Python data science packages

American Causal Inference Conference (ACIC) 2022

Participant

Berkeley, CA

May 2022 – May 2023

- Organized a data challenge team from Komodo Health; Inverse Probability Weighting Difference-in-Differences
- Corresponding author of resulting paper, published in *Observational Studies*
- See github.com/mepland/acic_causality_challenge_2022 for full details

Mount Sinai Health Hackathon

Participant

New York, NY

October 2019 – August 2023

- Prototyped an image based CNN classifier to read standard 12 lead EKGs in PyTorch
- Continued work with Dr. Subrat Das, MD published in *Current Problems in Cardiology*
- See github.com/mepland/mount_sinai_health_hackathon_ekg_img for full details

Relevant Technical Skills

- **Data Science:** Tree Based Models (FIGS, BDT, RF), Time Series Analysis, Survival Analysis, Neural Networks (NN)
- **Computing:** Python, SQL, DBT, R, Spark, scikit-learn, PyTorch, Pandas, Polars, Jupyter, C++, Tableau