Matthew McAnear

Curriculum Vitae
Friday 1st November, 2024

📍 Ann Arbor, MI
✉ mmcanear@umich.edu
⌨ mcanearm
🆔 0009-0006-3318-7387

Education

University of Michigan, College of Literature, Science, and the Arts   Expected: 2025
M.S. Applied Statistics

➤ Outstanding First Year Master's Student

University of Pennsylvania, School of Policy and Practice   2013-2014
M.S. Nonprofit/NGO Leadership

➤ Donald J.Deutsch Endowed Graduate Fellowship

Bucknell University   2009-2013
Dual B.A. Mathematics and Economics, magna cum laude

➤ National Merit Finalist Scholarship
➤ Dean's Scholarship
➤ Omicron Delta Epsilon Economics Honor Society

Research Experience

Graduate Student, University of Michigan   2022-present
Department of Statistics, Advised by Dr. Yang Chen

➤ Applied Gaussian process regression with novel contaminated noise extension to predict geomagnetic perturbations.
➤ Compiled the results of the model outputs into a manuscript, published in Data Science in Science.
➤ Currently, I am exploring the use of Kernelized Stein Discrepancy scores for recalibrating prediction intervals of arbitrary regressors when the marginal target distribution is known.

Research Assistant, Cooperative Institute for Great Lakes Research   2023-present
SEAS/NOAA, advised by Dr. Yi Hong and Dr. Dani Jones

➤ Fit multiple statistical models to predict Great Lakes water levels from climate forecasting reanalysis data (CFRS)
➤ Focused in particular on an autoregressive Gaussian process with hierarchical hyperparameters.
➤ Performed a model comparison to identify methods with best prediction interval coverage and accuracy.

Research Assistant

Teaching Experience

Graduate Student Instructor, University of Michigan   2024-present

➤ Taught lab sections for introductory statistics and data science courses
➤ Assist with grading of homeworks, quizzes, and exams
➤ Provided lecture support for lectures of 400+ students on staff of 20+ GSIs

Employment Experience

Point Predictive
Senior Data Scientist   2021-present

➤ Fit statistical models for predicting early loan default and chargeoff for auto lenders
➤ Write performant SQL for real-time and aggregate historical borrower statistics
➤ Load customer data into multiple database environments

➤ Implemented DynamoDB-backed solution for simple database lookups to improve AWS Lambda cold start times

Clear Capital

Senior Machine Learning Engineer 2019-2021

➤ Principal engineer for Clear Capital's AVM, a system that produces 150 million new predictions and 150GB of new data each week.
➤ Implemented a model-based error prediction procedure to create value ranges around AVM estimates
➤ Provide developer support and manage AWS infrastructure for the machine learning team, including launching and managing Amazon Quicksight dashboards for model analysis and validation.
➤ Designed and administer Redshift-based data warehouse to power machine learning models, ETL workflows, and ad-hoc analytical queries.
➤ Write, train, and deploy machine learning models on serverless and cloud infrastructure.

Data Scientist II 2017-2019

➤ Led team of five to simplify original AVM model, allowing more frequent, accurate, and cheaper builds, saving over $4,000 on compute costs each month and improving AVM performance from last place to industry leader in 6 months (based on absolute mean prediction error).
➤ Created lightweight web interfaces in R and Shiny for team usage in model validation and exploration.
➤ Designed and deployed the S3 storage and DynamoDB metadata layer for an internal photo application that manages 35+ terabytes of photos.
➤ Gained familiarity with a wider array of machine learning models, most notably random forests.

Data Scientist I 2015-2017

➤ Built an automated valuation model (AVM) on commodity hardware to predict home prices using R and PostgreSQL
➤ Implemented multi-model aggregation system for final prediction of house values on a monthly refresh cycle.
➤ Profiled and optimized R, Python, and SQL code for efficiency
➤ Built basic webservices in Python, Flask, and AWS Lambda for serving AVM model predictions and internal company data through a RESTful interface.

Seer Interactive

Data Scientist 2014-2015

➤ Designed and carried out web-based experiment on domain recognition using robust analytical methods, including multivariate regression and hierarchical modeling
➤ Wrote web-crawlers and multithreaded programs in Python for large-scale data collection jobs
➤ Utilized Linux-based computing resources in the cloud and on premises for batch processing, monthly reporting, and ad-hoc data analysis
➤ Automated repetitive reporting processes for large clients, including a hospital chain valued at $10 billion, by writing functions and scripts in R, Python, and Bash
➤ Wrote scripts and applications to extract data from various APIs and load them into databases stored both locally and in the cloud
➤ Supported analytics account managers and website tracking projects with external clients through project planning, KPI identification, and Google Analytics code implementation

Publications

1. Iong, D., McAnear, M., Qu, Y., Zou, S., Toth, G., and Chen, Y. (2024), "Sparse Variational Contaminated Noise Gaussian Process Regression with Applications in Geomagnetic Perturbations

Forecasting," Data Science in Science, Taylor & Francis, 3, 2383281. https://doi.org/10.1080/26941899.2024.2383281.