

Matthew McAnear

Senior Data Scientist

Resume

August 2023

 Ann Arbor, MI
 +1 423-277-0411
 mcanearm@gmail.com
 mcanearm

Professional Experience

- | | | |
|-------------|---|------------------------|
| 2021 - 2023 | Senior Data Scientist
Point Predictive | San Diego, CA (Remote) |
| | <ul style="list-style-type: none">➤ Fit statistical models for predicting early loan default and chargeoff for auto lenders in both object oriented and functional programming paradigms.➤ Created ETL pipelines for securely loading sensitive lender data (PPI) into NoSQL and RDBMS databases➤ Wrote performant Postgres stored procedures for real-time calculation of historical borrower data, returning business-critical aggregate statistics in sub-200ms query times.➤ Produced a PostgreSQL-based fuzzy matching probabilistic record linkage algorithm for retrieving prior credit applications of incoming borrowers, reducing query times and improving match rate each by 50%.➤ Designed and automated Snowflake queries to detect model distribution drift. | |
| 2020 - 2021 | Senior Machine Learning Engineer
Clear Capital | Reno, NV (Remote) |
| | <ul style="list-style-type: none">➤ Designed, launched, and administered Redshift data warehouse to power machine learning models, ETL workflows, business intelligence dashboards, and ad-hoc analytical queries.➤ Created Hadoop and Pyspark ETL pipelines for data engineering, model fitting, and batch prediction.➤ Spearheaded the creation of a quantile-regression error prediction procedure and accompanying Java implementation to recalibrate model prediction intervals.➤ Architected A/B testing framework in Quicksight and Jenkins for model iteration comparison➤ Built a Bayesian hierarchical model and accompanying Docker-based web-service to predict home complexity | |
| 2017 - 2019 | Data Scientist II
Clear Capital | Reno, NV |
| | <ul style="list-style-type: none">➤ Led team of five to streamline automated valuation model (AVM), reducing build time and costs by over 90%➤ Improved AVM performance from last place of 31 vendors to industry leader in 6 months, leading to annual revenues of over \$1M.➤ Designed and deployed an S3 and DynamoDB backed application that manages 35+ terabytes of MLS photos and their metadata.➤ Added temporal normalization to improve model performance and stability | |
| 2015 - 2017 | Data Scientist I
Clear Capital | Reno, NV |
| | <ul style="list-style-type: none">➤ Served as principal developer for Clear Capital's automated valuation model (AVM), a system that produces 150 million new predictions and 150GB of new data each week.➤ Built webservices in Python, Flask, and AWS Lambda for serving AVM model predictions and internal company data through a RESTful interface. | |
| 2014 - 2015 | Data Scientist
Seer Interactive | Philadelphia, PA |
| | <ul style="list-style-type: none">➤ Designed and carried out web-based experiment on domain recognition using multivariate hierarchical regression.➤ Supported analytics account managers and external clients through project planning and automated reporting. | |

Research Experience

- 2023 **Research Assistant** Statistics Department
University of Michigan
- Fit Gaussian process time-series models for prediction of solar weather events on Slurm cluster
 - Clean and analyze time-series data across multiple magnetometers
- 2014 **Research Assistant** School of Policy and Practice
University of Pennsylvania
- Support research on competition between charter and independent schools in Texas by gathering, cleaning, and analyzing raw data
 - Developed Stata program to merge, append, and manipulate multiple datasets to create research data for desired tests and analyses
 - Carried out analysis of children's survey responses for Comegys Elementary School on behalf of the Philadelphia Eagles
- 2012-2013 **Research Assistant** Bucknell Environmental Center
Bucknell University
- Investigated effects of social capital on health status using ACTION Health survey of 3300 respondents in the Central Susquehanna Region
 - Performed linear and logistic regressions and factor analyses using STATA and SPSS
 - Presented findings at Geisinger Research Symposium

Skills

Expertise

- Data Science, Machine Learning, Model Validation, Data Engineering, Databases, Dashboarding, A/B Testing, Time-Series Analysis, Visualization

Technologies

- R, Python, Linux, SQL, AWS, Snowflake, Docker, PostgreSQL, Looker, Slurm, Hadoop, Spark, REST, Flask, Git

Packages

- scikit-learn, PyMC, numpyro, jax, pandas, numpy, pyspark, PyTorch, sqlalchemy

Algorithms

- RandomForest, Regression, Clustering, XGBoost, MCMC sampling, Variational Inference, Bayesian Deep Learning, Hierarchical Modeling, Record Linkage, KNN, Clustering, MapReduce

Education

- Present **M.S., Applied Statistics** Ann Arbor, MI
University of Michigan
- 2014 **M.S., Nonprofit/NGO Leadership** Philadelphia, PA
University of Pennsylvania
- Donald J. Deutsch Endowed Graduate Fellowship
- 2013 **B.A., Mathematics & Economics** Lewisburg, PA
Bucknell University
- *Magna cum laude*
 - National Merit Finalist Scholarship
 - Dean's Scholarship
 - Omicron Delta Epsilon Economics Honor Society