

Matthew McAnear

Senior Data Scientist

Resume
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">➤ Developed statistical models to predict early default/chargeoff for auto lenders, reducing risk on millions of dollars of auto loans.➤ Constructed modular ETL pipelines to securely load sensitive lender data (PPI) into NoSQL and RDBMS databases, significantly reducing load times from over 8 hours to under 30 minutes.➤ Created high-performance Postgres stored procedures for real-time calculation of historical borrower data, delivering critical aggregate statistics with sub-200ms query times.➤ Designed a PostgreSQL-based fuzzy matching probabilistic record linkage algorithm to retrieve prior credit applications of incoming borrowers, resulting in 50% faster query times and improved match rates.➤ Automated Snowflake queries for model distribution drift detection, streamlining weekly efforts from 5 hours to just 10 minutes.	
2020 - 2021	Senior Machine Learning Engineer Clear Capital	Reno, NV (Remote)
	<ul style="list-style-type: none">➤ Launched and managed Redshift data warehouse, supporting ML models, ETL workflows, BI dashboards, and ad-hoc analytical queries.➤ Developed efficient Hadoop and Pyspark ETL pipelines for data engineering, model fitting, and batch prediction, resulting in an 80% reduction in build time.➤ Authored a quantile-regression based interval recalibration procedure, improving prediction interval coverage from 90% to the target of 68%.➤ Architected an A/B model testing framework, enabling parallel model development and standardizing model comparisons.➤ Replaced a heuristic home complexity scoring engine with a Bayesian model, achieving approximately 40% better performance and 99% lower costs.	
2017 - 2019	Data Scientist II Clear Capital	Reno, NV
	<ul style="list-style-type: none">➤ Led a team of five developers to optimize the automated valuation model (AVM), resulting in a reduction of build time and costs by over 90%.➤ Transformed AVM performance from last place among 31 vendors to a Top 3 model within just 6 months, contributing to additional annual revenues exceeding \$1 million.➤ Devised an event-driven S3 and DynamoDB application to manage 35+ terabytes of MLS photos and metadata with latencies consistently below 500ms.➤ Applied temporal normalization techniques to enhance model performance and stability, ensuring more reliable and accurate predictions.➤ Tailored the random-forest algorithm prediction process to enhance model precision, resulting in a 50% reduction in prediction outliers.	
2015 - 2017	Data Scientist I Clear Capital	Reno, NV
	<ul style="list-style-type: none">➤ Led a five person development team as the principal developer for Clear Capital's AVM, a system that generated 150 million weekly predictions and annual revenues exceeding \$5 million as of 2021.➤ Designed and implemented robust webservices using Python, Flask, and AWS Lambda to deliver AVM model predictions and internal company data in real-time through a seamless RESTful interface.	

- 2014 - 2015 **Data Scientist** Philadelphia, PA
Seer Interactive
- Conducted a web-based experiment on domain recognition using multivariate hierarchical regression and summarized results in a whitepaper.
 - Provided support to analytics account managers through programmatic report creation. This initiative cut over 80 hours off monthly reporting tasks.

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.
- 2014 **Research Assistant** School of Policy and Practice
University of Pennsylvania
- Supported research on competition between charter and independent schools in Texas by gathering, cleaning, and analyzing raw data.
 - 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.

Skills

Expertise

- Data Science, Machine Learning, Model Validation, Data Engineering, DBA, Query Optimization, 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