

# Matthew McAnear

Senior Data Scientist

Resume  
2024

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

## Professional Experience

2021 - 2023	<b>Senior Data Scientist</b> Point Predictive	San Diego, CA (Remote)
	<ul style="list-style-type: none"><li>➤ Developed statistical models to predict early default/chargeoff for auto lenders, reducing risk on millions of dollars of auto loans.</li><li>➤ 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.</li><li>➤ Created high-performance Postgres stored procedures for real-time calculation of historical borrower data, delivering critical aggregate statistics with sub-200ms query times.</li><li>➤ 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.</li><li>➤ Automated Snowflake queries for model distribution drift detection, streamlining weekly efforts from 5 hours to just 10 minutes.</li></ul>	
2020 - 2021	<b>Senior Machine Learning Engineer</b> Clear Capital	Reno, NV (Remote)
	<ul style="list-style-type: none"><li>➤ Launched and managed Redshift data warehouse, supporting ML models, ETL workflows, BI dashboards, and ad-hoc analytical queries.</li><li>➤ Developed efficient Hadoop and Pyspark ETL pipelines for data engineering, model fitting, and batch prediction, resulting in an 80% reduction in build time.</li><li>➤ Authored a quantile-regression based interval recalibration procedure, improving prediction interval coverage from 90% to the target of 68%.</li><li>➤ Architected an A/B model testing framework, enabling parallel model development and standardizing model comparisons.</li><li>➤ Replaced a heuristic home complexity scoring engine with a Bayesian model, achieving approximately 40% better performance and 99% lower costs.</li></ul>	
2017 - 2019	<b>Data Scientist II</b> Clear Capital	Reno, NV
	<ul style="list-style-type: none"><li>➤ Led a team of four senior developers to optimize the automated valuation model (AVM), resulting in a reduction of build time and costs by over 90%.</li><li>➤ 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.</li><li>➤ Devised an event-driven S3 and DynamoDB application to manage 35+ terabytes of MLS photos and metadata with latencies consistently below 500ms.</li><li>➤ Applied temporal normalization techniques to enhance model performance and stability, ensuring more reliable and accurate predictions.</li><li>➤ Tailored the random-forest algorithm prediction process to enhance model precision, resulting in a 50% reduction in prediction outliers.</li></ul>	
2015 - 2017	<b>Data Scientist I</b> Clear Capital	Reno, NV
	<ul style="list-style-type: none"><li>➤ Led a seven person development team of junior and senior analysts and data scientists 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.</li><li>➤ 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.</li></ul>	

2014 - 2015	<b>Data Scientist</b> Seer Interactive	Philadelphia, PA
	<ul style="list-style-type: none"><li>➤ Conducted a web-based experiment on domain recognition using multivariate hierarchical regression and summarized results in a whitepaper.</li><li>➤ Provided support to analytics account managers through programmatic report creation. This initiative cut over 80 hours off monthly reporting tasks.</li></ul>	

## Research Experience

2023	<b>Research Assistant</b> University of Michigan	Statistics Department
	<ul style="list-style-type: none"><li>➤ Fit Gaussian process time-series models for prediction of solar weather events on Slurm cluster.</li></ul>	
2014	<b>Research Assistant</b> University of Pennsylvania	School of Policy and Practice
	<ul style="list-style-type: none"><li>➤ Supported research on competition between charter and independent schools in Texas by gathering, cleaning, and analyzing raw data.</li><li>➤ Carried out analysis of children's survey responses for Comegys Elementary School on behalf of the Philadelphia Eagles.</li></ul>	
2012 - 2013	<b>Research Assistant</b> Bucknell University	Bucknell Environmental Center
	<ul style="list-style-type: none"><li>➤ Investigated effects of social capital on health status using ACTION Health survey of 3300 respondents in the Central Susquehanna Region.</li></ul>	

## 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	<b>M.S., Applied Statistics</b> University of Michigan	Ann Arbor, MI
2014	<b>M.S., Nonprofit/NGO Leadership</b> University of Pennsylvania	Philadelphia, PA
	<ul style="list-style-type: none"><li>➤ Donald J. Deutsch Endowed Graduate Fellowship</li></ul>	
2013	<b>B.A., Mathematics &amp; Economics</b> Bucknell University	Lewisburg, PA
	<ul style="list-style-type: none"><li>➤ <i>Magna cum laude</i></li><li>➤ National Merit Finalist Scholarship</li><li>➤ Dean's Scholarship</li><li>➤ Omicron Delta Epsilon Economics Honor Society</li></ul>	