

Matthew McAnear

Senior Data Scientist

Resume

July 2023

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

Summary

Experienced full stack data scientist adept at using Python, R, SQL, and AWS tools to create cost-effective and scalable solutions to business problems.

Professional Experience

- | | | |
|-------------|---|------------------------|
| 2021 - 2023 | Point Predictive
<i>Senior Data Scientist</i> <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.➤ Wrote ETL pipelines for securely loading sensitive lender data (PPI) into multiple database environments, improving load times from 12 hours to less than 10 minutes.➤ Implemented DynamoDB-backed remote key-value store to improve AWS Lambda cold start times and move large files out of application layer.➤ 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 by 50%.➤ Designed and automated Snowflake queries to detect model distribution drift, saving 5 hours per week in manual performance monitoring and proactively identifying model miscalibration for clients. | San Diego, CA (Remote) |
| 2020-2021 | Clear Capital
<i>Senior Machine Learning Engineer</i> <ul style="list-style-type: none">➤ Served as principal developer for Clear Capital's AVM, a system that produces 150 million new predictions and 150GB of new data each week.➤ Built a Bayesian hierarchical model and accompanying Docker-based webser-vice to predict home complexity➤ Spearheaded the creation of a quantile-regression error prediction proce-dure and accompanying Java implementation to recalibrate model prediction intervals.➤ Implemented Amazon Quicksight dashboards for model validation at multiple levels of analysis, including each individual home in the United States.➤ Designed, launched, and administered Redshift data warehouse to power machine learning models, ETL workflows, and ad-hoc analytical queries.➤ Created custom PySpark classes and ETL pipelines for model fitting and predic-tion. | Reno, NV |
| 2017-2019 | Clear Capital
<i>Data Scientist II</i> <ul style="list-style-type: none">➤ Led team of five to streamline original AVM model, reducing build time and costs by over 90%➤ Improved AVM performance from last place of 31 vendors to industry leader in 6 months (based on absolute mean prediction error), 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.➤ Promoted to Senior Machine Learning Engineer in 2020. | Reno, NV |

2015-2017	Clear Capital <i>Data Scientist I</i> <ul style="list-style-type: none">➤ Built an automated valuation model (AVM) cascade on commodity hardware to predict home prices using distributed, high performance R and PostgreSQL.➤ Created custom mixture methodology based on empirical and historical model performance for final value prediction.➤ Built webservice in Python, Flask, and AWS Lambda for serving AVM model predictions and internal company data through a RESTful interface.➤ Promoted to Data Scientist II in 2017.	Reno, NV
2014-2015	Seer Interactive <i>Data Scientist</i> <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.	Philadelphia, PA

Tools

Languages

- R, Python, bash, SQL

Technologies

- AWS, S3, RDS, Lambda, Cloudformation, DynamoDB, Redshift, Snowflake, Linux, Docker, PostgreSQL, Quicksight, Looker, Slurm, Spark, Sagemaker, REST, Flask, Git

Packages

- scikit-learn, PyMC, numpyro, jax, pandas, numpy, ggplot2, pyspark, PyTorch, sqlalchemy

Algorithms

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

Education

2022-Present	M.S., Applied Statistics University of Michigan	Ann Arbor, MI
2013-2014	M.S., Nonprofit/NGO Leadership University of Pennsylvania <ul style="list-style-type: none">➤ Donald J. Deutsch Endowed Graduate Fellowship	Philadelphia, PA
2009-2013	B.A., Mathematics & Economics Bucknell University <ul style="list-style-type: none">➤ <i>Magna cum laude</i>➤ National Merit Finalist Scholarship➤ Dean's Scholarship➤ Omicron Delta Epsilon Economics Honor Society	Lewisburg, PA