Matthew McAnear
Senior Data Scientist

Resume
July 2023

⚲ Ann Arbor, MI
☎ +1 423-277-0411
✉ mcanearm@gmail.com
in mcanearm

Professional Experience

2021 - 2023   **Senior Data Scientist**   San Diego, CA (Remote)
Point Predictive
- ➤ 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 PostgresSQL-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.

2020-2021   **Senior Machine Learning Engineer**   Reno, NV (Remote)
Clear Capital
- ➤ 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.
- ➤ Built a Bayesian hierarchical model and accompanying Docker-based webservice to predict home complexity

2017-2019   **Data Scientist II**   Reno, NV
Clear Capital
- ➤ 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 (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.

2015-2017   **Data Scientist I**   Reno, NV
Clear Capital
- ➤ 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**   Philadelphia, PA
Seer Interactive
- ➤ 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.

Skills

Languages
- ➤ R, Python, bash, SQL

Technologies
- ➤ AWS, S3, RDS, Lambda, Cloudformation, DynamoDB, Sagemaker, Redshift, Snowflake, Linux, Docker, PostgreSQL, Quicksight, Looker, Slurm, Hadoop, Spark, 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, MapReduce

Education

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