

Matthew McAnear

Ann Arbor, MI

+1 423-277-0411 | mcanearm@gmail.com | [mcanearm](https://www.linkedin.com/in/mcanearm)

Professional Experience

Senior Data Scientist

San Diego, CA (Remote)

POINT PREDICTIVE

2021 - 2023

- 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) for DynamoDB, Postgres, Redshift, and Snowflake data.
- 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%

Senior Machine Learning Engineer

Reno, NV

CLEAR CAPITAL

2020-2021

- Spearheaded the creation of a quantile-regression error prediction procedure and accompanying Java implementation to recalibrate model prediction intervals.
- Designed, launched, and administered Redshift data warehouse to power machine learning models, Hadoop and Pyspark ETL workflows, ad-hoc analytical queries, and model validation in Quicksight.
- Built a Bayesian hierarchical model and accompanying Docker-based webservice to predict home complexity

Data Scientist II

Reno, NV

CLEAR CAPITAL

2017-2019

- Led team of five to improve 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.

Data Scientist I

Reno, NV

CLEAR CAPITAL

2015-2017

- Built an automated valuation model (AVM) to predict home prices using distributed, high performance R and PostgreSQL.
- Built webservices in Python, Flask, and AWS Lambda for serving AVM model predictions and internal company data through a RESTful interface.

Data Scientist

Philadelphia, PA

SEER INTERACTIVE

2014-2015

- 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.

Tools

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, Bayesian Statistics, Hierarchical Modeling, Record Linkage, Clustering, MapReduce

Education

M.S., Applied Statistics

Ann Arbor, MI

UNIVERSITY OF MICHIGAN

Present

M.S., Nonprofit/NGO Leadership

Philadelphia, PA

UNIVERSITY OF PENNSYLVANIA

2014

B.A., Mathematics & Economics

Lewisburg, PA

BUCKNELL UNIVERSITY

2013