Paul Hein

Software Engineer · Big Data/ML Specialist

Redmond, WA | (+1) 520-289-9886 | paul.d.hein@gmail.com | github.com/pauldhein | linkedin.com/in/pauldhein

Education

University of Arizona

Aug. 2013 — June 2019

MS Computer Science — *Aug. 2017 — June 2019*

BS Computer Science / BA Mathematics — *Aug. 2013 — May 2017*

Experience

Rocket Mortgage

Sept. 2021 — Aug. 2023

Senior Machine Learning Engineer — *June 2023 — Aug. 2023*

- Lead a compute cost reduction of up to 95% for several data pipelines by training colleagues to use Apache Spark to improve data pipeline efficiency.
- Improved a lead delivery service by eliminating unreported errors and reducing remaining errors by 90% using AWS Lambda and SQS message timers.

Machine Learning Engineer — *Sept. 2021 — May 2023*

- Automated the validation of ETL updates by creating a dataset synthesizer system capable of creating nonprod safe replicas of production datasets.
- Enabled a marketing attribution model to be run daily by reducing compute costs by 40% through data modeling improvements with Apache Spark.
- Improved the technical maturity of a junior engineer through mentorship and pair-programming which lead to them receiving a promotion.
- Delivered a $400M increase in YoY mortgage loan volume by implementing an ML model for paid search optimization with AWS SageMaker and SciPy.
- Reduced ML model update rollout time from days to hours by aligning the data scientist experiment environment with the production environment.

ML4AI Laboratory

June 2016 — Sept. 2021

Research Software Engineer — *June 2019 — Sept. 2021*

- Contributed to the lab being awarded a new DARPA research grant by designing an ML model for generating Python source code from assembly code.
- Implemented a source code dataflow graph interchange format that became the standard for all external collaborators on a DARPA research grant.
- Provided technical mentorship to graduate students on proper utilization of PyTorch, NumPy, Scikit-Learn and experiment containerization via Docker.

Graduate Research Assistant — *May 2017 — June 2019*

- Achieved 6x training acceleration for sequence translation neural nets by implementing data parallelism in PyTorch over a distributed 8 GPU cluster.
- Improved the precision of an inter-sentence biomedical relation extraction model by 26 points using rigorous feature engineering and model selection.
- Designed a parallel hyperparameter grid search program in Python capable of tuning any Scikit-learn classifier on a distributed computing cluster.

Undergraduate Research Assistant — *June 2016 — May 2017*

- Created a corpora of musical patterns mined from jazz solos using a spatial pattern discovery algorithm for training an ML jazz solo generation model.
- Created a web application with Python, Flask, and D3.js capable of allowing an AI jazz generation model to record duets with a human musician.

Lunar Planetary Laboratory

April 2015 — June 2016

Student Programmer

- Assisted in developing a web application using Node.js that enabled scientists across the globe to view, create, and catalog spacecraft telemetry data.
- Assisted in designing a database ERD and implementing a SQL schema for pedigree tracking of data products originating from telemetry data.

Projects

BTD purchase predictor

July 2021 — Sept. 2021

- Developed machine learning classifiers to determine if a bank client would purchase a bank term deposit (BTD) using the UCI telemarketing dataset.
- Created a training pipeline with Python, Pandas, NumPy, Scikit-learn, class imbalance correction, and grid search to achieve an 89% AUC-ROC score.

Source code summarization

Jan. 2019 — May 2019

- Developed an encoder-decoder neural network using dyNet and NumPy to generate natural language summaries from Python function source code.
- Created a corpora of python functions and docstrings from the Python package index using NLTK, Word2Vec and regex for tokenization and encoding.

Skills

Engineering	Algorithm analysis • Software system design • Object oriented design • Test driven development • Functional programming
Data / ML	Supervised learning • Cluster algorithms • Data modeling • ETL • Data analysis • Feature engineering • Data visualization
Programming	Python • Java • JavaScript • C++ • Terraform • SQL • Bash • LaTeX • Spark • Pandas • NumPy • PyTorch • Scikit-learn
Technologies	Git • Docker • Kubernetes • Helm • AWS • CircleCI • Jira • GDB • Linux • Jupyter • LucidChart • Microsoft Office
Soft skills	Time management • Planning • Adaptability • Communication • Stress management • Teamwork • Problem solving