

Skills

Languages PYTHON ■ Java ■ Scala ■ R
ML/Data Science Pytorch ■ Jupyter ■ sklearn ■ numpy ■ scipy ■ matplotlib ■ ggplot ■ Shiny ■ tidyr
Devops/Infra Kubernetes ■ Docker ■ AWS ■ Jenkins ■ Terraform ■ Ansible ■ Packer
App Dev Flask ■ React ■ NodeJS

Work Experience

Roam Analytics **ML Platform Engineering Lead.**
2018-2020 Built autoscaling, GPU-enabled Kubernetes cluster on Amazon EKS, using custom AMLs for healthcare-grade security ■ Trained, deployed, and served hundreds of containerized machine learning models and workloads ■ Generated massive healthcare knowledge graph (billions of edges) using Airflow ■ Built NLP-powered medical text search and annotation applications backed by a healthcare knowledge graph on a React/Flask/Elasticsearch stack ■ Wrote infrastructure as code on AWS using Terraform, Ansible and Packer ■ Implemented continuous integration and deployment pipelines on Jenkins, using tools including pylint, mypy, Docker, and Ansible

Wealthfront **Senior Data Engineer.**
2016-2018 Wrote 30+ Spark pipelines for big data processing ■ Sped up batch event writes to database by 2x compared to the Hibernate ORM ■ Implemented Spark cluster autoscaling + optimized parititioning, reducing daily runtime of all batch jobs from 24+ hours to 15 hours ■ Reduced complexity of managing and debugging our job-dependency graph by over twofold, by sorting dependency graph with a topological sort and removing redundant dependencies. ■ Built low-latency key-value store based on RocksDB ■ Wrote automated data quality checks and real-time monitoring of compute clusters

Yelp **Search Quality Intern.**
Summer 2015 Improved search results for misspelled queries, by adding Kneser-Ney smoothing to the language models underlying Bayesian query categorization

Citadel **Software Intern.**
Summer 2014 Developed and deployed statistical model of trade-execution latency in C++ and R

UCLA **Applied Mathematics Research.**
Summer 2013 Time series modeling of crime data with stochastic differential equations

Education

Cornell University **Master of Engineering, Computer Science, GPA 4.0.**
2020-2021 Deep Probabilistic and Generative Models; Large Scale Machine Learning; Reinforcement Learning, Natural Language Processing; Computational Linguistics II; Networks and Markets; Game Theory; Behavior and Information Technology

Harvey Mudd **Bachelor of Science, Mathematics, GPA 3.7.**
College Intro to CS; Principles of Computer Science; Data Structures and Program Development; Algorithms, Computational Biology; Bayesian Statistics; Time Series; Statistical Linear Models; Abstract Algebra; Partial Differential Equations; Intermediate Linear Algebra; Algebraic Geometry

Other Skills

Languages English - Native ■ Japanese - Intermediate