

Matt Motoki

Python Data Processing & ML Pipelines · Kaggle Grandmaster

Remote | mattmotoki.github.io | [Kaggle](#) | [GitHub](#) | [Google Scholar](#)

Data scientist and Python developer with over a decade of experience building efficient, accurate data-processing and machine-learning pipelines. Comfortable owning the full path from raw data to validated output: ingestion, cleaning, transformation, feature computation, and integration with existing systems. Optimized **terabyte-scale processing** with incremental updates, vectorization, and careful choice of data structures. Kaggle Competitions Grandmaster with 13 national and international awards — a record built on fast, correct data pipelines as much as on modeling.

EXPERIENCE

Shifted Energy — Staff Data Scientist | Honolulu, HI / Remote · Aug 2021 – Jun 2025

Led production data and machine-learning pipelines for thousands of connected devices worldwide — ingesting, validating, and scoring live time-series data with monitored, reproducible jobs backed by SQL databases.

H2O.ai — Senior Data Scientist | Mountain View, CA / Remote · Jun 2020 – Aug 2021

Built algorithms and product features for **enterprise AutoML software** in Python, with heavy emphasis on efficient data handling, correctness, and performance across customer datasets of widely varying size and quality.

Iterable — Senior Data Scientist | San Francisco, CA · Jun 2019 – May 2020

Developed machine-learning and feature-engineering workflows over **terabyte-scale event data**, using incremental (streaming-style) updates to avoid full recomputation and keep processing fast and cheap.

Independent — Data Science Consultant | Honolulu, HI · Jul 2017 – Jun 2019

Built predictive models, data pipelines, analyses, and dashboards for a remote-sensing startup, academic researchers, and local startups.

HiVet Solutions — Data Scientist | Honolulu, HI · Mar 2016 – May 2017

Developed data-driven methods for workforce planning, hiring, and compensation analysis with HR and Finance teams.

SELECTED COMPETITION RESULTS

- **NASA Airport Throughput Prediction Challenge**, 2025 — **1st Place · \$30,000** — efficient processing of large historical operations data
- **Digitizing Utilities Prize** (U.S. DOE), 2023 — **Phase 1 Winner · \$75,000** — analytics application turning messy operational data into decisions
- **ASHRAE Great Energy Predictor III**, 2019 — **1st Place · \$10,000** — large-scale tabular prediction on Kaggle (3,600+ teams)

Also: ten further awards in prediction, forecasting, and optimization — \$350,000+ total across 13 awards (2017–2026).

EDUCATION

- **Doctoral Research, Electrical & Computer Engineering** — University of Washington · 2022 – 2025 — machine learning, game theory & optimization for large-scale prediction
- **M.S., Electrical Engineering** — University of Hawai'i at Mānoa · Dec 2015
- **B.S., Mathematics** — University of Hawai'i at Mānoa · Aug 2014

CORE SKILLS

Python (pandas, NumPy, scikit-learn, PyTorch) · SQL & PostgreSQL · Data structures & algorithmic optimization · Performance profiling, vectorization & incremental computation · ETL & reproducible data pipelines · Data validation & accuracy checks · Machine learning · Data visualization (matplotlib, Plotly) · Git & code review