Jack Stehn

Machine Learning Engineer

San Francisco, California • jack@ste.hn • www.linkedin.com/in/jack-stehn • github.com/jstehn • ste.hn

PROFESSIONAL SUMMARY

Machine Learning Engineer specializing in building and deploying production ML systems end-to-end. I bridge the gap between 'notebook data science' and production-ready systems by bringing strong software engineering practices to predictive modeling and scalable MLOps infrastructure. I focus on ensuring models are explainable, maintainable, and directly impactful at scale.

EXPERIENCE

Data Scientist (Lead: ML, MLOps, Data Engineering) - Ed Pioneers Fellow

Sep 2024 — Oct 2025

Caliber Public Schools

- **Predictive ML & Risk Modeling:** Developed and deployed explainable ML models (Logistic Regression, Deep Learning) to predict staff turnover. Engineered a 'Risk Tolerance' configuration allowing non-technical leadership to adjust precision/recall thresholds based on quarterly hiring capacity.
- **Modern Data Stack Architecture:** Architected and built a scalable ML platform on GCP. Orchestrated ELT pipelines using Dagster, dbt, and dlt to ensure clean, reliable feature data for model training.
- **Engineering Maturity (ROI):** Built a custom automated data quality framework to catch inconsistencies between systems, reducing manual consistency checks from months of annual work to seconds.
- **Strategic Leadership (Solo Data Lead):** Owned the full data lifecycle (DS, DE, ML) as the sole data professional, partnering directly with C-suite to identify high-leverage ML applications.

Data Scientist (Production ML, NLP, MLOps)

Aug 2021 — Feb 2023

SetSail

- **Production ML (Revenue):** Developed and deployed production ML models for Propensity Scoring (deal closure probability) and Churn Modeling. Leveraged NLP on unstructured email metadata to identify sales signals.
- **Business Impact:** Contributed to ML-driven product enhancements that achieved 33% faster ramp times, 16% higher revenue, and 15x ROI for customers.
- **Pipeline Architecture (AWS):** Led a critical overhaul of the AWS data infrastructure (S3, Athena, EMR). Implemented 'SQL Push-down' strategies and asynchronous DAGs, reducing data processing latency by 75% and scaling to multi-terabyte datasets.
- **Scalable Data Modeling:** Architected scalable Star Schema data models and optimized ETL/ELT processes, ensuring data readiness for LLM integration.
- **Engineering Best Practices:** Championed the adoption of CI/CD pipelines (GitHub Actions), unit testing (pytest), and Agile methodologies within the data science team.
- **Causal Analysis:** Performed deep causal inference studies to isolate specific sales behaviors that drive outcomes, influencing the product roadmap.

Data Science Research Team Lead

Sep 2020 — May 2021

UC Berkeley School of Public Health

- **Leadership:** Led data science components for mixed-methods studies on equity and public health. Managed a team of undergraduates.
- **Unstructured Data:** Analyzed diverse unstructured and non-traditional datasets (qualitative interviews, geospatial data, text corpora) requiring novel data processing approaches.

EDUCATION

University of California, Berkeley

Aug 2019 — May 2021

Bachelor of Arts in Data Science (Domain Emphasis: Quantitative Social Science)

GPA: 4.00/4.00

Highest Distinction (Summa cum laude). Outstanding Data Science Undergraduate Award (Top of Class).