KD BARTHOLOMEW

✉ bartholomewkd@gmail.com · 💼 linkedin.com/in/kd-bartholomew-34002495 · ⬤ kdbartholomew.github.io

Machine Learning Engineer and AI Developer with 8 years' experience in probability, statistics, and linear algebra. Expert in Python and building scalable, production-grade deep learning systems that deliver real-world impact.

Education

Master of Science in Data Science	**San Francisco, CA**
University of San Francisco	*Jul 2024 – Jun 2025*
Bachelor of Science in Data Science	**San Francisco, CA**
University of San Francisco	*Aug 2022 – Jun 2024*

Experience

Data Scientist	**San Francisco, CA**
Queer Life Space	*Feb 2025 - Present*

- Engineered and deployed **A/B testing pipeline** that **increased donations 15%** and implemented **HIPAA-compliant ETL pipeline to build a dynamic pricing model** that **boosted client retention by 10%**.
- Developed and deployed an analytics dashboard, **accelerating leadership decision-making by 25%**.

Data Engineer	**San Francisco, CA**
The Nature Conservancy (Intern)	*Sep 2024 - Present*

- Built a **scalable AWS geospatial pipeline** to process **600+ GB of satellite imagery**, achieving over **97% water detection accuracy**, **cutting processing time by 50%**, and **saving $225K in year one** compared to traditional river monitoring.

Machine Learning Engineer	**San Francisco, CA**
University of San Francisco (Researcher)	*Aug 2023 - Jun 2024*

- Developed a **custom U-Net neural network** achieving **98% pixel accuracy** for automated cell counting, **saving 60 hours per month and $36K in annual labor costs**.

Math Instructor	**San Francisco, CA**
Independent	*Jan 2017 - Aug 2022*

- Taught AP and **college-level calculus, differential equations, linear algebra, probability, and statistics to 100+ students,** adapting lessons to diverse learning styles and emphasizing real-world applications.

Projects

Weight Loss AI Assistant

- Built a production-grade **agentic RAG system** combining **fine-tuned LLMs, CV models, and vector DBs** for automated food logging and adaptive coaching.
- Developed and deployed a **scalable multi-agent pipeline** using **LangChain, Docker, and GCP**, with a Postgres backend enabling real-time agent coordination.

Distributed Sentiment Analysis Pipeline

- Built a distributed **Apache Spark pipeline** on **GCP and MongoDB Atlas** to process **500K+ daily rows** of financial and social media text.
- Deployed **Hugging Face models** for real-time sentiment analysis, converting unstructured data into a **queryable NLP system**.

Recommendation System for Growth Curves

- Developed a robust recommendation system combining Funk SVD matrix factorization with a custom neural network, **improving prediction accuracy by 40% despite 30%+ missing data.**

Metropolis-Hastings for Cryptography

- Engineered and optimized Metropolis-Hastings, a **probabilistic sampling algorithm**, **in a novel way, enabling convergence on large-scale cryptographic inputs that traditional methods couldn't handle.**

Publications & Awards

Founder Award – USF Innovation Summit	**Apr 2025**

- Built **AI Hire**, a **privacy-first job matching platform** featuring a **fine-tuned Sentence Transformer** and an **edge-deployed inbox scraper** for automated application tracking; awarded for **full-stack AI innovation**, including **LLM fine-tuning**, **edge AI**, **cloud architecture**, and **user-centric product design**.

Best Overall Nationally – American Statistical Association	**Dec 2017**

- Led impactful statistical analysis of Seattle police data, producing actionable insights adopted citywide to improve community safety; **awarded Best Overall Nationally** and **published in *AMSTATNEWS***, the official American Statistical Association newsletter.

Technical Skills

Core Competencies: Probability, Statistical Analysis, Machine Learning, Linear Algebra
Programming Languages: Python, R, SQL, NoSQL, Bash, Zsh
Data Science & ML: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, Hugging Face, Pandas, NumPy, CUDA
Big Data & Cloud: AWS, GCP, Apache Spark (PySpark), Airflow, PostgreSQL, Snowflake, MySQL, MongoDB
Deployment & DevOps: Docker, Kubernetes, CI/CD (GitHub Actions, Jenkins), Terraform, MLflow, MetaFlow
Visualization & Reporting: Matplotlib, Seaborn, Plotly, Geopandas, Flask, FastAPI, Power BI, Streamlit, Tableau