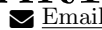# KD BARTHOLOMEW

(408) 499-6130 ✉ Email in LinkedIn ◯ GitHub

San Francisco, CA

Data Scientist / Machine Learning Engineer with over 4 years of experience coding in Python to build scalable pipelines, ML models, and impactful data solutions. Experienced in leveraging cloud services and 2 years working with deep learning frameworks.

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 (Consultant) — *Feb 2025 - Present*
- Engineered and deployed an A/B testing pipeline that **boosted donation conversions by 15%** through user behavior analysis, and built a dynamic pricing model that promotes nonprofit financial stability while **increasing client engagement by 10%**.
- Designed an interactive dashboard for real-time trend monitoring, **accelerating data-driven decisions by 20%**.

Data Engineer — **San Francisco, CA**
The Nature Conservancy (Intern) — *Sep 2024 – Present*
- Engineered a **scalable AWS geospatial pipeline** processing **600+ GB of satellite imagery** with spectral bands and NIR differencing, achieving **97%+ accuracy** in surface water detection and **reducing processing time by 50%**.
- Leveraged **remote sensing** and **cloud-native data pipelines** to advance **conservation technology**, enhancing water resource monitoring and supporting **biodiversity preservation** across California ecosystems.

Machine Learning Engineer — **San Francisco, CA**
University of San Francisco (Researcher) — *Aug 2023 - Jun 2024*
- Developed a custom U-Net neural network achieving **95% pixel accuracy** for automated cell counting, **saving 60 hours/month and $36K in annual labor costs**.
- Preprocessed microscopy image datasets from the Schroeder Lab at USF using data augmentation and normalization techniques to optimize the dataset for deep learning. Optimized CUDA training, **reducing processing time by 45%** and **improving generalization by 15%**.

Math Instructor — **San Francisco, CA**
Independent — *Jan 2017 – Aug 2022*
- Taught AP and **college-level calculus, differential equations, linear algebra, probability, and statistics.**
- Adapted lessons to student learning styles, emphasizing connections between concepts and real-world applications to deepen understanding and critical thinking.

Projects

Weight Loss AI Assistant
- **Fine-tuned and prompt-engineered CV and LLM models** for automated food logging and personalized guidance, leveraging **retrieval-augmented generation (RAG)** and **vector databases** for context-aware coaching.
- Designed and built a **custom multi-agent system** with **LangChain agent frameworks**, orchestrated with **Docker** and **AWS/GCP** for scalable, modular AI workflows.
- Architected, deployed, and managed a **Postgres database on GCP** to support scalable agent collaboration, user data persistence, and real-time retrieval for **production-ready AI pipelines.**

Distributed Sentiment Analysis Pipeline
- **Built Apache Spark** pipeline on **Google Cloud Storage (GCS)** and **MongoDB Atlas** to process and analyze approximately **500K daily rows (2.5 GB/week)** of financial news headlines and YouTube comments.
- Converted **unstructured text** into a structured, queryable database for scalable **NLP** and **sentiment analysis**.
- Deployed **Hugging Face** sentiment models to reveal media-public sentiment alignment via **cloud-native** large-scale processing.

Recommendation System for Growth Curves
- Engineered a **recommendation system** leveraging **Funk SVD matrix factorization** to impute missing data, reveal **latent growth patterns**, and drive **personalized, data-driven insights**.
- Designed and trained a **custom neural network** that **boosted performance by 40%** despite **30%+ missing data**, enhancing system accuracy and resilience.

Metropolis-Hastings for Cryptography
- Engineered and optimized Metropolis-Hastings, a **probabilistic sampling algorithm**, for cryptographic applications **in a novel way, enabling convergence on large-scale cryptographic inputs that traditional methods couldn't handle.**
- Incorporated log scaling and change point detection into the algorithm's scoring function, **improving computational efficiency and scalability for high-dimensional cryptographic problem-solving.**

Publications & Awards

Founder Award – USF Innovation Summit — **Apr 2025**
- Founded and built **AI Hire**, a privacy-first AI job matching and tracking platform featuring a **fine-tuned Sentence Transformer model** for resume-job matching and a **locally deployed (edge AI) inbox scraper** for automated job tracking.
- Received **Founder Award for full-stack innovation, spanning LLM fine-tuning, edge AI deployment, cloud infrastructure planning,** and product strategy to scale intelligent, user-centric hiring tools.

Best Overall Nationally – American Statistical Association — **Dec 2017**
- Led statistical analysis of Seattle police data for the **AMSTATNEWS & ThisIsStatistics Police Data Challenge**, delivering **actionable insights implemented by the Seattle Police Department** to enhance community safety.
- Awarded **Best Overall Nationally** and **published findings in the official newsletter of the American Statistical Association**.

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