KD Bartholomew

(408) 499-6130 Email LinkedIn GitHub
San Francisco, CA

Data scientist with over 5 years of experience coding in Python to build scalable pipelines, ML models, and impactful data solutions. Experienced in leveraging cloud services and 3 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 remote sensing 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%.**

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**.
- Applied normalization and data augmentation during preprocessing and 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 to 100+ students,** adapting lessons to diverse learning styles and emphasizing real-world applications to deepen understanding and critical thinking.

Projects

Weight Loss AI Assistant

- **Fine-tuned CV and LLM models** with **RAG** and **vector databases** for automated food logging and personalized coaching.
- Built a **custom multi-agent system** using **LangChain**, **Docker**, and **AWS/GCP** for scalable AI workflows.
- **Architected and managed a Postgres database on GCP** enabling scalable agent collaboration and real-time user data retrieval in production.

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

- 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**, for cryptographic applications **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**

- Founded and built **AI Hire**, an end-to-end, 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: ASA & 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 AMSTATNEWS**, 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