KD BARTHOLOMEW *(they/them)*

(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

MS in Data Science Jul 2024 – Jun 2025
University of San Francisco *San Francisco, CA*

BS in Data Science Aug 2022 – Jun 2024
University of San Francisco, Cum Laude *San Francisco, CA*

PROFESSIONAL EXPERIENCE

Data Scientist & Engineer Feb 2025 – Present
Queer Life Space (Consultant) *San Francisco, CA*

- Engineered and deployed an A/B testing pipeline for the organization's website. Optimized donation strategies, resulting in a 15% increase in donations through conversion rate optimization and user behavior analysis.
- Developed a dynamic sliding scale pricing model using linear regression, ensuring financial sustainability while maintaining accessibility for clients. Improved client affordability and financial health of the organization, leading to a 10% increase in client engagement.
- Designed and deployed a real-time interactive dashboard to monitor pricing trends, enabling stakeholders to make data-driven adjustments aligned with nonprofit goals; improved decision-making speed by 20%.

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

- Built a scalable geospatial data pipeline using AWS S3, Planet Labs APIs, and NDWI spectral analysis for remote sensing, handling hundreds of gigabytes of geospatial data. Reduced data processing time by 50%, enabling faster decision-making.
- Implemented computer vision algorithms (Otsu's thresholding, Canny edge detection) to automate water body classification and river flow analysis.
- Experimented with and implemented time-series analysis models (LSTM, ARIMA, Prophet, XGBoost) to predict low-flow events, enhancing proactive resource management.

Data Scientist Aug 2023 – Jun 2024
University of San Francisco (Researcher) *San Francisco, CA*

- Engineered and implemented a custom U-Net architecture for automated cell counting in microscopy images, achieving an average pixel accuracy of 95%. Improved efficiency by saving researchers approximately 60 hours/month, leading to estimated annual savings of $36,000 in 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 model training with CUDA, reducing processing time by 45%. Fine-tuned hyperparameters and applied regularization techniques, leading to a 15% improvement in model generalization on unseen data.

PROJECTS

Recommendation System for Growth Curves with Funk SVD Matrix Factorization (Python)

- Applied Funk SVD matrix factorization to impute missing data and uncover latent growth patterns.
- Developed a custom sequential neural network that enhanced performance by 40% despite 30% missing data.

Metropolis-Hastings for Cryptography (Python)

- Engineered and optimized Metropolis-Hastings (MH) algorithms 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.

Super-positioning in Neural Networks (Python, PyTorch)

- Developed toy feed-forward ReLU-based models to explore neural network super-positioning and pattern representation.
- Analyzed noisy synthetic datasets to prototype, visualize, and interpret neural network super-positioning.

PUBLICATIONS & AWARDS

AMSTATNEWS: American Statistical Association
ThisIsStatistics: Police Data Challenge - Best Overall Nationally

- Led a project to wrangle, analyze, and present Seattle police data, delivering actionable insights that were implemented by the Seattle Police Department to enhance community safety.
- Won "Best Overall" nationally and published findings in the American Statistical Association's journal, showcasing the measurable impact of data-driven approaches on public safety initiatives.

TECHNICAL SKILLS

Core Competencies: Probability Theory, Statistical Analysis, Machine Learning, Linear Algebra
Programming Languages: Python, R, SQL, NoSQL, Java
Data Science & Machine Learning: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, Hugging Face Pipelines, Pandas, NumPy, CUDA
Big Data & Cloud: AWS S3, Google Cloud Platform, Apache Spark (PySpark), Apache Airflow, PostgreSQL, Snowflake, MySQL, MongoDB
Tools & Environments: Git, GitHub Actions, VS Code, Docker, Bash/Zsh
Visualization & Reporting: Matplotlib, Seaborn, Plotly, Geopandas, Flask, FastAPI, Power BI, Streamlit