Caleb Tucker Dame

www.github.com/calebdame
www.linkedin.com/in/caleb-dame/
calebdame@gmail.com
calebdame.com

Experience

Senior Data Scientist – BlastPoint **Dec. 2025 – Present**
Data Scientist – BlastPoint **Oct. 2023 – Nov 2025**
Remote Work

- Design predictive model products for strategic product marketing, churn prediction, collections optimization, or other additional use cases to be for credit unions and utility partners, used to drive adoption rates and measure retention strategies.
- Develop and implement privacy-affirming ensemble federated prediction models, dynamically aggregating insights from internal data to develop adaptable, industry-wide product models while protecting original client data privacy.
- Scale team throughput at a startup by building robust internal tooling scripts to standardize processes, promote best practices, increase velocity through tooling, monitor model performance, and expose model interpretability to end users.

Data Scientist – TransUnion **Feb. 2022 – Oct 2023**
Remote Work

- Built and shipped final products to the TransUnion TruValidate product suite for identity fraud detection and IP Address risk models using boosted trees (LightGBM/XGBoost) and designed custom unsupervised drift detection metrics.
- Researched fraud rings by leveraging identity databases to create identity graphs for feature creation and demonstrated the viability of Graph Neural Networks as complex feature extractors that can identify synthetic identity or fraud propagation.
- Led cross-team collaborations with new subsidiaries, integrating new data assets to mutually enrich model performance.

Data Engineer – M Science LLC **Jun. 2021 – Mar 2022**
Murray, UT

- Designed and maintained 15+ ETL pipelines in Databricks, Airflow, and Snowflake, processing 3.5 TB of transformed data daily for financial analysts.
- Automated text anomaly detection using clustering algorithms on NLP-derived features, increasing data purity by 10–15% and enabling real-time cohort visualization in Tableau dashboards.

Machine Learning Researcher – Brigham Young University, Economics **Apr 2020 – Jul 2020**
Provo, UT

- Researched feature engineering techniques to help find matches across census years, interpreting the census sheets and rows as a graph to find correlations in family & locality dynamics to finally reduce record-linking error by an additional 80%.

Relevant Projects

- **HOA CC&R RAG Pipeline** – A containerized PDF ingestion service that performs OCR and text extraction on large PDFs, semantic chunking, and a FAISS-based vector search to inform a "Question Answering" RAG pipeline regarding HOA Covenant documents.
- **Symbolic Music Transformer (Custom Attention)** – Improved performance and training speed of a symbolic music generation transformer by adding graph-based attention biases (e.g., circle-of-fifths, timing, instrument) to an existing transformer architecture.
- **RL Policy Pretraining via Genetic Algorithms** – Investigated evolutionary pretraining for deep RL Continuous Control problems to reduce variance and early catastrophic failures by warm-starting policy discovery using a populationg of evolving candidates.

Education

Johns Hopkins University – M.S. Artificial Intelligence **Aug 2022 – Present**
Brigham Young University – B.S. Applied and Computational Mathematics & Economics **Aug 2017 – April 2021**
Dean's List Honoree (2018, 2020)

Technical Skills

Programming & Data Science

- **Python:** Numpy, Scipy, Matplotlib, Scikit-Learn, Pytorch, Keras, Tensorflow, StatsModels, Pandas, NLTK
- **SQL & Databases:** SQLite, PostgreSQL, NoSQL
- **Big Data & Cloud:** Databricks, Apache Spark (*PySpark*), GCP, AWS (*Sagemaker, Lambda, EC2, DynamoDB, Redshift*)
- **Other Languages**: JavaScript, C++, Unix Shell, HTML, STATA

Development and Tooling

- **Environments and Versioning:** Git, Docker, Conda
- **Data Visualization:** Tableau, Grafana, Matplotlib, Seaborn
- **Workflow Automation & MLOps:** Flask, FastAPI, Cron, AB Testing, MLFlow, Airflow