Tan Dai Ngo

250-661-9530 | Victoria, BC | ngotandai95@gmail.com | linkedin.com/in/ntdai95/ | ntdai95.github.io

Technical Skills

- **Programming:** Python, SQL, Java, MATLAB
- **Machine Learning:** Regression, tree-based models, feature engineering, cross-validation, optimization
- **Data & ML Tools:** Pandas, NumPy, PyTorch, scikit-learn, XGBoost, CatBoost
- **Systems & Deployment:** SQLite, Cassandra, Neo4j, Kafka, Docker, FastAPI, AWS (EC2)

Education

Master of Engineering in Applied Data Science — **September 2025 - August 2026 (Expected)**
- GPA: 9.0 / 9.0 — University of Victoria
- Relevant Coursework: Optimization for Machine Learning, Data Analysis and Pattern Recognition, Systems for Massive Datasets, Information Theory, Algorithms and Data Models

Master of Science in Computer Science — University of Chicago

Bachelor of Science in Economics — University of Washington

Experience

Software Developer | T-Mobile (via BeaconFire Inc.) — *Bellevue, WA, USA* | **June 2022 - April 2024**
- Collaborated with a team of 5 to develop enterprise microservices for the Roaming Business System (Spring Boot, Java Stream API), integrating with Jenkins CI/CD for deployments and Splunk for log monitoring.
- Designed and optimized Cassandra database schemas for partner and workflow management microservices, handling 100k+ records while maintaining query execution time under 3 seconds.
- Automated weekly reprocessing of 50+ roaming service tests with Kafka pipelines, cutting manual effort by 80%.

Projects

Market Sentiment and Volatility Intelligence System (Team of 4) — Algorithms and Data Models Course
- Architected an end-to-end pipeline using FinBERT and PyTorch to process 15k+ tweets; implemented batched inference and automated daily sentiment aggregation for Top 10 tech stocks (e.g., NVDA, AAPL).
- Built a hybrid infrastructure using Neo4j (Cypher) and SQLite to map influencer-stock networks; engineered complex queries to isolate high-impact market events based on engagement metrics.
- Applied GARCH(1,1) and Granger Causality to quantify social sentiment's impact on market volatility; identified sentiment as a lead indicator for volatility clusters rather than raw price returns.

End-to-End Machine Learning System for Multi-Output Fuel Blending — Shell.ai Hackathon 2025
- Deployed a production-ready machine learning system using FastAPI, Docker, and AWS EC2 to serve real-time multi-output regression predictions for 10 chemical blend properties.
- Benchmarked XGBoost vs. CatBoost using 5-fold cross-validation, selecting CatBoost after achieving lower error (MAPE 0.64 vs. 1.29) and productionizing the best-performing model.
- Engineered robust feature and inference pipelines using weighted property aggregation, entropy-based mixture metrics, and serialized preprocessing to ensure consistent production inference.

Handwritten Digit Recognition System — Optimization for Machine Learning Course
- Developed a Multi-Class SVM and Softmax pipeline using a custom-built ML-BFGS optimizer, achieving 97.65% accuracy with faster convergence than standard SGD.
- Leveraged HOG and PCA to reduce 784-pixel dimensions to 50 latent components, resulting in a 10.2% accuracy lift and 40% faster training latency.
- Optimized regularization parameters (μ, C) and evaluated performance via 10*10 confusion matrices to ensure robust generalization on 10,000 test samples.