Tan Dai Ngo

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

Technical Skills

- **Programming & Data Engineering:** Python (Numpy, Pandas), SQL, Java, MATLAB, Cassandra, Neo4j, Kafka
- **Machine Learning & Data Science:** supervised and unsupervised learning, multi-output regression, gradient boosted decision trees (XGBoost, CatBoost), scikit-learn, feature engineering, cross-validation, model evaluation
- **Deployment:** Docker, FastAPI, AWS (EC2), CI/CD (Jenkins), model serialization, log monitoring (Splunk)

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

Sentiment Stock Analysis (Team of 4)   Algorithms and Data Models Project
- Architected a multiple data source system using Neo4j for graphical influencer mapping and SQLite for relational time series storage, ingesting X (Twitter) API streams and historical market data.
- Implemented batched FinBERT inference pipeline to generate sentiment signals achieving up to 59% directional accuracy in predicting next day returns during volatile periods.
- Validated predictive relationships between social signals and returns via Granger causality and modeled volatility periods using GARCH(1,1).

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 Project
- Developed a multiclass 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.