

DHRUV BHUTANI

☎ 437-982-5258 ✉ dhruvb.2604@gmail.com 🔗 linkedin.com/dhruvbhutani 🌐 dhruvbhutani.me

TECHNICAL SKILLS

Programming: Python (pandas, NumPy, PySpark), SQL **Machine Learning:** Scikit-learn, XGBoost, PyTorch, TensorFlow **Data Engineering:** ETL/ELT, Apache Spark, Databricks, Airflow **Cloud & Databases:** AWS (S3, Lambda, SageMaker), GCP Vertex AI, MySQL, DynamoDB **Deployment & Tools:** Docker, Git, CI/CD, FastAPI, Streamlit

EXPERIENCE

XE.com

Sept 2024 – Dec 2024

Machine Learning Engineer Intern

- Built a geospatial ML pipeline on **AWS Lambda** with vectorized Python services; processed **200K locations/day** and cut feature engineering time by **25%**.
- Improved fraud-detection performance by training and tuning **XGBoost** models with Bayesian search; raised AUC from 0.78 to **0.85**.
- Created and deployed **retrieval-based GPT-4 prototypes** via **FastAPI** endpoints to support fraud-detection demos used by multiple teams.

Mistplay

May 2024 – Aug 2024

Machine Learning / Data Science Intern

- Assisted in building a recommendation system using **matrix factorization** and **Transformer-based NLP models (Hugging Face, LangChain)** to enhance personalization.
- Developed and optimized **data pipelines** on **Databricks (PySpark, SQL)** for large-scale transaction processing and model training.
- Supported deployment of ML models as **FastAPI** microservices, enabling near real-time predictions in production.

XE.com

Sept 2023 – Dec 2023

Applied ML Intern — Fraud Detection

- Built hybrid ML pipelines using **PyTorch**, **XGBoost**, and **AdaBoost** to detect anomalous patterns across **100K transactions/week**.
- Set up cross-cloud ML infrastructure on **AWS SageMaker, S3, and DynamoDB**, reducing model deployment time from **days to under 1 day**.

Teranet Inc.

Jan 2023 – Apr 2023

ML Engineer Intern — Real Estate Analytics

- Trained ensemble models (**Random Forest**, **XGBoost**) for property valuation; improved prediction accuracy by **10%**.
- Optimized **ETL pipelines** with **Pandas/Scikit-learn** and **Docker**; reduced processing time from **2 hours to 40 minutes**.

BCN Labs, Bain & Company

May 2022 – Aug 2022

AI Research Intern — Forecasting Systems

- Developed a no-code **forecasting platform** with **TensorFlow**, **Streamlit**, and pretrained **Transformer** models; enabled non-technical staff to deploy forecasts.
- Benchmarked deep learning architectures (**ANNs**, **CNN-LSTMs**, **Transformers**) on structured time-series data; identified models with **10% lower error**.

PROJECTS

- **Real-Time Stock Market Data Streaming Pipeline:** Built a real-time stock market pipeline with **Apache Kafka** (Python producers/consumers), streaming JSON data from CSVs. Applied on-the-fly **pandas** transformations in **Jupyter Notebooks** to simulate live market conditions at one-second intervals.
- **F1 Data Platform – Automated ETL with Airflow & AWS S3:** Orchestrated daily ETL pipelines from the **FastF1 API** into **PostgreSQL** using **Airflow DAGs**; automated backups to **AWS S3**. Containerized workflows with **Docker**, added task-level data quality checks and S3 verification for scalable analytics and streaming readiness.

EDUCATION

The University Of Waterloo

Bachelor of Science in Computer Science

- **Relevant Coursework:** Machine Learning, Computational Vision, Database Management & Statistics in CS

May 2025
Waterloo, ON