

BHARGAV KUMAR NATH

Data Scientist & Applied Scientist

+44 7563 131573 · bhargavkumarnath@gmail.com · [LinkedIn](#) · [GitHub](#) · [Portfolio](#)

EDUCATION

MSc Data Science and Analytics · Merit (2:1) · University of Leeds, UK Sep 2024 - Dec 2025
BTech Computer Science and Engineering · First Class Honours · Assam Don Bosco University, India Sep 2020 - Sep 2024

TECHNICAL SKILLS

Quantitative Analysis: Causal Inference (X-Learner), Experimentation (A/B Testing, Bandits), Hypothesis Testing, SHAP

Modeling & Science: PyTorch, Scikit-Learn, LightGBM, XGBoost, Time-Series Forecasting, Gaussian HMM, HDBSCAN

Data Computing: Python, R, SQL, Polars, DuckDB, Pandas, NumPy, Pandera, Docker, Git, GCP

PROFESSIONAL EXPERIENCE

Data Analyst Intern · Airports Authority of India, Regional HQ Jul 2023 - Aug 2023

- Engineered automated data mapping workflows for enterprise SAP ERP systems, reconciling over 1,000 distributed asset profiles into a standardized tracking schema.
- Built an automated ETL pipeline to replace manual logging, giving regional operations teams real-time visibility into hardware availability.

Data Engineering Intern · Indian Institute of Technology, Guwahati Jul 2022 - Aug 2022

- Designed a heuristic constraint-satisfaction algorithm to resolve multi-variable academic scheduling conflicts, balancing hard institutional limits with faculty availability.
- Architected an ACID-compliant MySQL relational schema across 15 interdependent modules, applying targeted clustering indexes to minimize high-frequency query latency.

Junior ML & Data Analyst · M/S Sanjog Trading, India Jul 2020 - Nov 2021

- Wrote Python ingestion pipelines (Pandas, NumPy) to clean high-volume transactional logs, normalizing messy operational records into structured database tables.
- Implemented statistical time-series forecasting models to predict seasonal demand swings, directly informing reorder schedules to prevent inventory stockouts.

TECHNICAL PROJECTS

Dynamic Causal Experimentation Engine · Python, LightGBM, Polars, X-Learner, LinUCB [\[GitHub\]](#) [\[Live\]](#)

- Architected an X-Learner causal pipeline on 14M records, identifying that a naive A/B test's **+59% lift** masked a **-\$0.05/user loss**, and reversed it to a **+\$0.09 profit/user (+280%)** using a profit-aware LinUCB bandit policy.
- Deployed the model for **Real-Time Bidding (RTB)** by distilling the 5-model LightGBM ensemble into a surrogate **Decision Tree**, slashing inference latency from **120ms to ~45μs** while retaining **>95% economic value**.

Andria Systems: Institutional Quant Research Platform · DuckDB, Polars, HDBSCAN, Gaussian HMM [\[GitHub\]](#) [\[Live\]](#)

- Engineered an out-of-core pipeline using DuckDB and HDBSCAN to process **116M SEC 13F filings** within a **16 GB** memory limit, clustering **4,000+** asset managers into **4 behavioral archetypes**.
- Modeled regime-conditioned equity signals with a **Gaussian Hidden Markov Model** and a custom event-study backtester, achieving a statistically significant **1.847** out-of-sample **Sharpe ratio** across **10 walk-forward folds**, net of T+1 execution costs.

FinSight-Alpha: Rigorous NLP Evaluation Agent · LangGraph, RAGAS, Qdrant, Cross-Encoder [\[GitHub\]](#) [\[Live\]](#)

- Engineered an autonomous document analysis agent with a self-correcting hallucination detection loop via a **6-node LangGraph** state machine, achieving a **0.91 RAGAS Faithfulness score** against a **0.71** baseline.
- Eliminated false-positive retrievals via a hybrid **BM25** and **Qdrant** dense vector search pipeline fused with **Reciprocal Rank Fusion** and cross-encoder reranking, lifting retrieval precision to **91%** and F1 to **89%**.

PUBLICATIONS

'The Evolution of AI' · 'Beyond the Hill: Global Optimisation' · 'ESG in the Age of AI' (Leeds Financial Insights)