BHARGAV KUMAR NATH

Data Scientist

+44 7563 131573 · bhargavkumarnathh@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, Power BI, DAX

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

- Analyzed 1,053+ IT asset records using enterprise dashboards to track hardware lifecycle, deployment patterns, and maintenance data, identifying trends in asset utilization and replacement requirements.
- Evaluated procurement and vendor transaction data from the Government e-Marketplace (GeM) and documented SAP ERP data workflows across HR, Finance, and Procurement systems to understand data relationships and reporting processes.

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

- Designed a normalized MySQL database schema with optimized indexing to manage user, scheduling, and transactional data across 15+ application modules, ensuring efficient data storage and retrieval.
- Developed automated scheduling logic in Node.js to handle complex date constraints including holidays, leave periods, and scheduling conflicts, while analyzing user research data to guide application improvements.

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

- Built Python ETL pipelines using pandas, Polars, and NumPy to process 800K+ sales records, automating data ingestion, transformation, and preparation for analytics workflows.
- Developed and evaluated time-series forecasting models using scikit-learn, XGBoost, statsmodels, and PyTorch, comparing performance with RMSE and custom metrics to support inventory planning.
- Implemented an anomaly detection pipeline using Isolation Forest and deployed Streamlit and PowerBi dashboards to visualize forecasts, sales trends, and irregular transaction patterns.

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 $\sim$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)