Anirudh Malik

Data Scientist

+91 9058176356 anirudhforjobs@gmail.com
LinkedIn GitHub Portfolio Delhi, India

Summary

Data Scientist with strong foundations in machine-learning, statistical analysis, and data engineering. Experienced in developing scalable ML models, optimizing data pipelines, and delivering actionable business insights. Adept at exploratory data analysis, data wrangling, and visual storytelling. Proven ability to deploy solutions in production using tools like Docker, AWS, and MLflow.

Experience

Research Associate **IIT Delhi**
Delhi, India **04/2025 – Present**

- Simulated Higgs boson production via $gg \to H \to \tau^+\tau^-$ using MadGraph5 + Pythia8 + Delphes
- Conducted exploratory data analysis (EDA) on CMS MiniAOD datasets for gen vs reco tau leptons
- Optimized statistical cuts and applied data mining techniques to improve signal purity
- Produced reproducible analysis using Jupyter, ROOT, and Git-based pipelines
- Collaborating with 3 physicists and reporting progress at weekly lab reviews

AI/ML Specialist ***Imagenators***
Noida **05/2025 – 07/2025**

- Built and deployed a containerized RAG system using FastAPI, LangChain, FAISS, and Gemini 1.5
- Developed REST endpoints for PDF ingestion, data wrangling, and real-time LLM-based question answering
- Benchmarked generative models vs classical NLP on retrieval accuracy
- Managed data pipeline development with PostgreSQL, Docker Compose, and continuous integration tests

Trainee Developer ***Dev Group***
New Delhi **02/2025 – 05/2025**

- Designed a CO_2 emissions prediction model using regression techniques with Flask + MLflow
- Conducted data wrangling, EDA, and created dashboards for cross-team reporting
- Contributed to internal analytics tools using Python REST APIs and Pickle model serialization

Graduate Researcher **University of Sheffield**
Sheffield, UK **09/2023 – 09/2024**

- Developed PICNN classifiers for $Z \to e^+e^-$ using ATLAS MC data
- Used Optimal Transport to match AF3 and Geant4 simulations

Undergraduate Researcher **University of Delhi**
Delhi, India **09/2022 – 05/2023**

- Simulated surface plasmon resonance via DDSCAT and analyzed dielectric properties of $AgNO_3$ nanomaterials

Education

- **MSc in Particle Physics**, University of Sheffield, UK
  2023 – 2024 First Division
- **BSc in Physical Science**, University of Delhi, India
  2020 – 2023 Distinction
- **Senior Secondary (ISC)**, Mount Carmel School
  2018 – 2020 94%

Skills

- Python, SQL, C++, Bash
- Data wrangling, Data mining, EDA
- Statistical analysis, Hypothesis testing
- PyTorch, TensorFlow, scikit-learn
- Data pipeline dev, MLflow, Airflow
- REST, FastAPI, Flask, Django
- LangChain, Ollama, ChromaDB
- Pandas, NumPy, PostgreSQL, MongoDB
- Matplotlib, Seaborn
- AWS (EC2, S3, Lambda)
- Git, GitHub, Docker, Kubernetes
- ROOT, Geant4, MadGraph5
- Apache Spark, Big Data Analytics

Certifications

- Data Science, ML, DL, NLP
- Mastering SQL & Analytics

Self-Initiated Projects

Asteroid Discovery with NASA **2021**
- Participated in the International Astronomical Search Collaboration; discovered a provisional asteroid (awaiting MPC confirmation).

Hohmann Transfer Orbit Simulator **2021**
- Built an interactive GUI to demonstrate fuel-efficient orbit transitions in the Solar System.

Fourier Drawing using Rotating Vectors **2022**
- Created a visualization of Fourier Series via animated vector rotations.

Retail Sales Forecasting Pipeline **2023**
- Built a full-stack ML pipeline using Airflow, Flask, and MLFlow; boosted forecast accuracy by 20

Fraud Detection System with MLOps **2024**
- Designed a scalable ensemble-based fraud detection system using Apache Spark and AWS Lambda.

Languages

English — Professional **Hindi** — Native