RAJAT GUPTA

DATA SCIENTIST

CONTACT

+33-783454133

rajatgupta116@gmail.com

Geneva, Switzerland
(Open to Remote)

https://www.linkedin.com/in/dr-rajat-gupta-a4a77391/

github.com/rajat116

SKILLS

Programming & Tools

- Python, C++, Bash, SQL, Git, Jupyter, Docker, Conda, Advanced Excel, Makefile, FastAPI, Pre-commit

Machine Learning & Deep Learning

- Regression, Classification, Clustering, PCA, Anomaly Detection, Isolation forest, Autoencoders (VAE), CNNs, Generative Models, XGBoost, Neural Networks

Frameworks & Libraries

- Scikit-learn, PyTorch, TensorFlow, Keras, Pandas, NumPy, Matplotlib, MLflow, Evidently

Data & Analysis

- HDF5, CSV, JSON, Parquet; Hypothesis Testing, Confidence Intervals, Uncertainty Estimation

Platforms & Environments

- Google Colab, Jupyter, GitHub, CUDA (GPU), HTCondor, SLURM, Docker, Singularity, MLflow, Mage, XGBoost, ML Pipelines, Terraform, GitHub Actions, Streamlit

Soft skills:

- Storytelling with data
- Solid interpersonal and collaborative skills

AWARDS & HONORS

- Breakthrough Prize in Fundamental Physics (2025) : – awarded as part of the ATLAS Collaboration at CERN.
- Ranked among top 3% of researchers globally (AD Scientific Index, 2023)
- Gold Medalist (1st rank) – M.Sc. Physics, Panjab University

LANGUAGES

- English (Fluent)
- French (Basics)

Data Scientist with a PhD in Physics and expertise in machine learning, data analysis, and scientific computing. Experienced in building robust models for complex, noisy datasets and applying model distillation techniques to compress and accelerate deep learning workflows. Skilled in Python, SQL, and modern ML frameworks, with a strong track record of driving data-driven insights and solutions across cross-functional teams.

PROJECT HIGHLIGHTS

GitHub Anomaly Detection (End-to-End MLOps Project)

- Built a real-time anomaly detection system using Isolation Forest and engineered GitHub activity features
- Orchestrated ETL, drift detection, and retraining with Airflow; tracked experiments and models via MLflow
- Deployed FastAPI inference service with Docker and CI/CD; provisioned MLflow via Terraform (IaC)
- Implemented monitoring with Streamlit dashboards, Evidently reports, and Slack/email alerts for drift and anomalies

WORK EXPERIENCE

Data Scientist **2022 - PRESENT**

ATLAS@LHC (CERN and University of Pittsburgh)

- Leading an international team of 10 researchers on an ongoing machine learning project for rare event detection; coordinating workflows, model development, and future publication.
- Developing and evaluating gradient boosting and neural network models for classification in high-noise data environments.
- Developing model distillation workflows to convert deep learning outputs into lightweight decision tree models for faster inference.
- Working with hardware engineers to tailor machine learning algorithms for low-latency applications, including real-time environments and FPGA-based systems.
- Partnering with an industry lab to build data-driven compression solutions for edge memory platforms, integrating ML techniques for efficient storage and retrieval.
- Mentoring one PhD and two undergraduate students in applied machine learning, data analysis and workflowas, and research communication.

Data Analyst **2016 - 2022**

CMS@LHC (CERN and Panjab University)

- Analyzed large-scale experimental datasets to extract structured patterns and validate data-driven hypotheses.
- Coordinated multi-year projects involving data quality checks, statistical comparisons, and reporting of key metrics.
- Co-authored multiple peer-reviewed publications; led communication and documentation for collaborative research.

EDUCATION

Ph.D. in Particle Physics **April 2022**

Panjab University)