CONTACT

- 📞 +33-783454133
- ✉️ rajatgupta116@gmail.com
- 📍 Berlin, Germany (Open to Remote)
- 🔗 https://www.linkedin.com/in/dr-rajat-gupta-a4a77391/
- 🐙 github.com/rajat116
- ✉️ www.rajat-gupta.dev

SKILLS

Programming & Tools

Python, SQL, Bash, DuckDB, Pandas, NumPy, Polars

Machine Learning

XGBoost, LightGBM, CatBoost, Scikit-learn, Feature engineering, Time-series forecasting (ARIMA, Prophet), Anomaly detection (Isolation Forest, Autoencoders/VAEs), Model explainability (SHAP), A/B testing

Deep Learning & LLMs

PyTorch, TensorFlow, Keras, CNNs, VAEs, Generative models, Sentence-Transformers, Vector search (Qdrant), RAG pipelines

MLOps & Deployment

MLflow, Docker, Airflow, FastAPI, GitHub Actions (CI/CD), Monitoring & drift detection (Evidently), Model packaging & API deployment

Cloud

AWS (S3, EC2), Terraform

Data Engineering & Analytics

Parquet, SQL analytics, ETL pipelines, Data validation, Statistical testing, Performance optimization

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)

LANGUAGES

- English (Fluent)
- French (Basics)

RAJAT GUPTA
DATA SCIENTIST/ ML ENGINEER

Data Scientist & ML Engineer with strong end-to-end experience building production-ready ML systems for risk modelling, anomaly detection, and behaviour prediction. I work with large, high-frequency datasets and have delivered models into real-time and batch environments using Python, SQL, MLflow, Airflow, and FastAPI. Experienced with tree-based models, deep learning, and LLM/RAG pipelines, and comfortable collaborating with cross-functional teams to turn complex data into measurable product impact.

PROJECT HIGHLIGHTS

Customer Credit Risk & Behaviour Scoring

Built an end-to-end PD/LGD + anomaly detection pipeline to solve the need for accurate borrower default prediction, enabling real-time millisecond scoring via FastAPI + Streamlit with SHAP explainability.
Tech: Python, SQL, XGBoost, LightGBM, FastAPI, MLflow.

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

Designed an Airflow pipeline to address the challenge of detecting abnormal user/repo behaviour in millions of GitHub events, delivering fully automated daily inference and real-time drift alerts.
Tech: Airflow, Docker, AWS, MLflow, Evidently, Python.

LLM Analytics Assistant (RAG + Vector Search)

Developed a Qdrant-based RAG system to solve slow, manual analytics over large structured datasets, enabling sub-second semantic querying through a Streamlit interface.
Tech: Qdrant, Sentence-Transformers, FastAPI, LLM APIs, Streamlit.

WORK EXPERIENCE

Data Scientist   2022 - PRESENT
ATLAS@LHC (CERN and University of Pittsburgh)

- Designed and deployed ML models for anomaly detection and real-time classification on 1B+ high-frequency events, building end-to-end pipelines for feature engineering, training, optimisation, and validation.
- Built model-distillation workflows and co-developed low-latency ML models with engineers for hardware-constrained environments, reducing inference latency to sub-3 µs.
- Developed scalable data pipelines and automated analysis workflows, improving model throughput and reducing manual effort for the team. Created data-driven compression solutions using tree models and autoencoders, improving storage & retrieval efficiency for high-throughput pipelines.
- Mentored PhD and undergraduate students on applied ML, data processing, and reproducible analytics.

Data Analyst   2016 - 2022
CMS@LHC (CERN and Panjab University)

- Processed and analysed large-scale structured event data to identify patterns, anomalies, and key performance signals using statistical modelling and Python-based analytics.
- Built reproducible data pipelines for cleaning, validation, and feature extraction across multi-TB datasets, improving team throughput and analysis reliability.
- Collaborated with physicists, engineers, and analysts to translate findings into clear insights adopted in workflow improvements.

EDUCATION

Ph.D. in Particle Physics   April 2022
(Panjab University)