Aditya Gupta

[redacted] | [redacted] | [redacted] | github.com/Adityag0243

Profile Summary

Data Science AI student at IIT Guwahati with strong foundations in Machine Learning, Statistical Inference, Time Series, and Optimization. Experienced in building end-to-end data science workflows—from data ingestion and feature engineering to deploying models on cloud infrastructure. Worked on structuring unorganized data into machine-readable formats and delivering practical, production-ready solutions. Strong problem-solving skills with 750+ competitive programming problems and a track record of taking projects from idea to deployment.

Technical Skills

Languages: Python, SQL, R, C++, Java
ML & Statistics: Scikit-learn, XGBoost, Random Forest, Regression Analysis, Hypothesis Testing, Statistical Inferencing, Feature Engineering, EDA, Model Evaluation (AUC-ROC, F1), Deep learning, NLP, Transformer Architecture
AI & GenAI: RAG Systems, LangChain, Qdrant (Vector DB), Ollama (Llama 3.1), Prompt Engineering
Data & Visualization: Pandas, NumPy
Tools & Cloud [Basic Hands-On]: Docker, GCP (Compute Engine), Git, FastAPI, Streamlit, Linux (Bash)
Core Concepts: DSA, OOP, DBMS, Optimization, Econometrics, Recommender Systems

Experience

AI Engineer Intern Jan 2026 – March 2026
LeapX (Physics Wallah) *Onsite*

- One of the core member who built **Souli**, an AI wellness agent – designed and implemented the full **data ingestion pipeline**: chunking, embedding, and indexing psychiatric transcripts into a **Qdrant** vector store for semantic retrieval.
- Translated an unstructured wellness domain (100+ hrs of youtube videoe's learning) into a structured **Energy Node Schema**, enabling deterministic query-to-framework mapping via cosine similarity – moving from an open-ended, ambiguous problem to a machine-queryable knowledge system.
- Optimized **LLM inference pipelines** for production chatbots using **KV Cache** and conversation history management, reducing response latency by **20%** with measurable system-level impact.
- Engineered **prompt injection defenses** using tool binding and refined structured prompt templates; resolved **10+ bugs** to ensure a stable, deployment-ready codebase across multiple production LLM projects.

Projects

Customer Churn Prediction | *Python, Scikit-learn, XGBoost, Pandas, Matplotlib, Seaborn*

- Built an end-to-end classification pipeline on **10,000+ customer records**, achieving **AUC-ROC of 0.88** by benchmarking **Random Forest and XGBoost** with cross-validated hyperparameter tuning.
- Performed systematic **EDA and feature engineering** to surface primary churn drivers; analyzed **feature importances** to translate model outputs into interpretable, retention-focused business recommendations.

Real-Time Voice AI Pipeline | *Python, LiveKit, Whisper, Ollama, Qdrant, FastAPI, GCP*

- Engineered a real-time voice agent on a **GCP L4 GPU VM** integrating **Whisper** (STT), **Llama 3.1** via Ollama (LLM inference), and **Qdrant RAG** – orchestrated end-to-end with **LiveKit** for low-latency audio streaming.
- Built a **FastAPI** integration layer handling LiveKit token generation and routing between the voice pipeline and mobile app backend, enabling seamless real-time voice interactions.

Education

Indian Institute of Technology Guwahati Guwahati, Assam
B.S. in Data Science & Artificial Intelligence (Online Degree) | *CPI: **8.1/10.0*** *Sep 2023 – Present*

- *Relevant Coursework*: Machine Learning Fundamentals (AB), Statistical Inferencing (AB), Basic Econometrics (AB), Time Series Analysis & Forecasting (BB), Recommender Systems (BB), Deep Learning Essentials (BB), Optimization (BB)

Jawahar Navodaya Vidyalaya Azamgarh, Uttar Pradesh
Secondary & Higher Secondary | *GPA: **9.5/10.0*** | *Physics, Mathematics, CS* *Sep 2015 – March 2022*

Achievements & Leadership

Winner – PW Institute of Innovation Hackathon | *1st of 50 teams* Jan 2025

- Shipped a D2C brand prototype end-to-end under time pressure – website frontend, Figma UI, and a complete business model.

Competitive Programming & Mentorship

- Codeforces **Pupil** (1368) | CodeChef **3 Star** (1642) | **750+** problems solved. Led DSA sessions for **15+ juniors** as Tech Club Co-Lead.