Anshuman Maurya

Ghaziabad, India | +91-8318463051 | anshumanmaurya776@gmail.com
LinkedIn | GitHub | LeetCode | YouTube

Profile Summary

Aspiring Data Scientist and AI/ML Engineer with hands-on experience in end-to-end Machine Learning, Deep Learning, and Generative AI pipelines. Skilled in Python, SQL, LangChain, and MLOps workflows. Builds production-grade AI solutions with strong focus on explainability, reproducibility, and data-driven decision making.

Education

Ajay Kumar Garg Engineering College 2023 – 2027
B.Tech in Information Technology
CGPA: 7.8 / 10.0

Projects

Superstore Intelligence Engine — ML-Driven Retail Analytics Platform GitHub | Live Demo
Enterprise-grade MLOps platform for retail intelligence, combining synthetic data generation, predictive modeling, and explainable AI.

- Designed end-to-end **MLOps pipeline** including ETL, automated EDA, and model training with reproducible workflows.
- Generated **30,000+ synthetic transactions** using CTGAN to handle class imbalance and improve model robustness.
- Built dual prediction engine: **classification (profitability) + regression (profit forecasting)** using XGBoost and Scikit-Learn.
- Implemented centralized **feature store** to prevent training-serving skew across pipeline stages.
- Integrated **SHAP explainability** and deployed interactive **Streamlit dashboard** for real-time inference.

AI-Powered SQL Q&A System (SQL-RAG) GitHub | Live Demo
Deterministic NLP-to-SQL system that converts natural language queries into executable SQL for real-time analytics.

- Built **LangChain SQLDatabaseChain** pipeline to translate user queries into optimized MySQL queries.
- Eliminated hallucinations by enforcing **database-grounded responses** from a live Railway-hosted MySQL instance.
- Implemented result parsing and validation logic for structured outputs and reliable analytics.
- Deployed cloud-based **Streamlit application** enabling real-time inventory insights via natural language.

FinSolve Role-Based RAG Chatbot GitHub | Live Demo
Secure enterprise AI assistant using Retrieval-Augmented Generation (RAG) with Role-Based Access Control (RBAC).

- Designed **RAG pipeline** using OpenAI embeddings and ChromaDB for semantic search across department-specific knowledge.
- Implemented **RBAC system** ensuring zero cross-department data leakage (Finance, HR, Engineering, C-level roles).
- Built scalable backend using **FastAPI** and interactive frontend with **Streamlit**.
- Integrated **GPT-4o-mini** for context-aware responses with role-filtered document retrieval.
- Deployed full-stack system with real-time chat, session memory, and secure API-based architecture.

Technical Skills

ML / DL	PyTorch, TensorFlow, Keras, Scikit-Learn, XGBoost, LightGBM, CNN, Transfer Learning, Data Augmentation
Agentic AI	LangChain, LangGraph, RAG, Tool Calling, Multi-Agent Systems, Groq LLM, FAISS, Prompt Engineering
NLP	SpaCy, NLTK, Hugging Face Transformers, TF-IDF, Word2Vec, NER, Cosine Similarity
Computer Vision	OpenCV, Image Preprocessing, Face Detection, Real-time Inference
Data & EDA	Pandas, NumPy, Matplotlib, Seaborn, Jupyter Notebook, Google Colab
Deployment	Streamlit, Flask, REST APIs
Languages	Python, C++, SQL
Tools	Git, GitHub, VS Code, Kaggle

Certifications & Achievements

- GeeksforGeeks — Data Science Project Based Learning
- Solved 250+ DSA problems on LeetCode with 200+ day streak.
- AI-focused YouTube creator sharing ML and GenAI content.