Ritik Pandey

📞 +91-8318480847 ✉ panditritik260@gmail.com ⚫ github.com/ritikpandey33

Education

Indian Institute of Information Technology Ranchi — Ranchi, Jharkhand
B.Tech in Electronics and Communication (CGPA: 7.6) — *2022 – 2026*

SRD Senior Secondary School, Mau — Mau
Class XII (CBSE) — 80.2% — *2019 – 2020*

Experience

Open Source Contributor

LlamaIndex

- Contributed 4 merged pull requests (500+ lines of code) to LlamaIndex, an open-source data framework for LLM applications with 44K+ GitHub stars.
- Fixed critical bug in ConfluenceReader preventing on-premise deployments from fetching child pages by implementing conditional logic to handle cloud vs on-premise instances with comprehensive test coverage.
- Resolved initialization crash in QdrantVectorStore when using async-only clients with hybrid search enabled, ensuring backward compatibility while adding support for legacy sparse vector formats.
- Extended Fireworks AI integration by adding support for additional models, improving framework compatibility.
- Implemented async support for AmazonKnowledgeBasesRetriever using aioboto3 for non-blocking retrieval.
- Collaborated with core maintainers through iterative code reviews.

Projects

Chat with Database | *Python, LangChain, Groq, Streamlit, SQL* | GitHub
- Engineered a Generative AI application leveraging LangChain's Agent and Groq-hosted LLaMA 3/Gemma LLMs to translate natural language queries into executable SQL
- Designed and implemented a dynamic Streamlit UI featuring session memory, role-based CSV upload, and interactive Altair data visualizations
- Integrated both SQLite and MySQL database connections within a unified interface for real-time querying and data interaction

Multi-Modal RAG Educational Platform | *Python, LangChain, LangGraph, ChromaDB, FastAPI, Streamlit, OpenAI* | GitHub
- Architected a multi-modal RAG system processing PDFs, YouTube transcripts, and web content with ChromaDB vector storage and configurable LLM providers via factory pattern
- Implemented diversity-aware retrieval algorithms to ensure balanced context assembly from text and video sources for educational queries
- Built a FastAPI backend with async endpoints for Q&A, MCQ generation, essay evaluation, and knowledge graph extraction using LangGraph validation workflows
- Developed a Streamlit interface for automated study guide generation featuring validated MCQs, essay prompts with rubrics, and interactive NetworkX knowledge graphs

Credit Card Fraud Detection System | *Python, Scikit-learn, XGBoost, Random Forest, ML* | GitHub
- Developed a high-performance fraud detection ML model achieving 99.12% AUC on a dataset of 284,000 transactions, with 85.4% recall and 82% precision
- Deployed a production-ready Random Forest model with advanced regularization, reducing overfitting to a minimal 2% performance gap between training and test sets
- Built a preprocessing pipeline to handle extreme class imbalance (0.17% fraud rate) using stratified sampling, feature scaling, and encoding techniques

Technical Skills

Programming Languages: Python, C++, SQL
ML/Deep Learning Frameworks: TensorFlow, Keras, Scikit-learn, PyTorch, LangChain, Hugging Face
Tools & Platforms: Git, Docker, Streamlit, FastAPI, Flask, Jupyter, Google Colab, MySQL, SQLite
Areas of Expertise: Machine Learning, Deep Learning, NLP, Generative AI, RAG, Data Preprocessing, Feature Engineering, Model Evaluation, MLOps
Data Visualization: Matplotlib, Seaborn, Power BI, Excel

Achievements

Competitive Programming

- **LeetCode:** Ranked 278 out of 25,000+ participants in LeetCode Weekly Contest 451.
- **LeetCode Rating:** Achieved a 1729 rating on LeetCode after solving 300+ algorithm challenges.
- **Codeforces Specialist:** Attained a peak rating of **1455** on Codeforces.