Abhay Yadav

+91-9696994723 | abhay131203@gmail.com | linkedin.com/in/abhay-ryad | github.com/abhayryad | leetcode.com/abhayryad

Technical Skills

Core Skills: Python, SQL, Java, Machine Learning, Deep Learning, Generative AI (RAG & LLMs), Agentic AI
ML & AI Tools: NumPy, Pandas, Scikit-learn, TensorFlow, LangChain, LangGraph, LangSmith, MLflow
Backend & Deployment: Docker, FastAPI, Hugging Face Spaces, AWS (EC2, Bedrock workflows)
Databases: MySQL, MongoDB
Analytics & Visualization: Power BI, MS Excel

Education

Vellore Institute of Technology (VIT) — Bhopal, India
B.Tech in Computer Science Engineering — CGPA: 8.47 — *Sep 2022 – Present*

Lucknow Public School — Lucknow, India
Class XII — 84.6% — *2022*

St. Mary's Inter College — Lucknow, India
Class X — 92% — *2020*

Projects

Security System – ML Pipeline & Deployment

Python, Docker, AWS EC2, MLflow, FastAPI, MongoDB

- **ML Pipeline Design**: Designed and implemented a **6-stage end-to-end ML pipeline** (data ingestion, preprocessing, training, validation, inference, deployment) processing **10K+ records per run**.
- **Deployment & Automation**: Containerized services using **Docker** and deployed on **AWS EC2**, reducing manual deployment steps from **8+ to 2** and cutting deployment time from **48 hours to under 3 hours** (**94% reduction**).
- **Reproducibility & Tracking**: Tracked **30+ experiments** using **MLflow** with data and model versioning, enabling reproducible training and consistent evaluation across multiple runs.
- **API Performance**: Developed a **FastAPI**-based inference service handling **20+ requests/min** during testing, with average response latency under **250 ms**.

lexSQL

Python, LangChain, Hugging Face, Gradio, RAG

- **System Design**: Designed and implemented a **schema-aware RAG system**, embedding **100% of database schema metadata** for accurate SQL generation.
- **Pipeline Engineering**: Built a modular architecture with **5 core components** enabling scalable and maintainable NL-to-SQL generation.
- **Performance & Cost**: Processed **100+ user queries** with consistent SQL correctness, achieving **sub-300ms retrieval latency** and **zero-cost deployment** using open-source LLMs.

Engine Sound Classifier

Python, TensorFlow, CNN, Librosa

- **Deep Learning Model**: Developed a CNN-based audio classifier for predictive maintenance, achieving **94% accuracy** across **4 engine conditions**.
- **Feature Engineering**: Extracted **40-dimensional MFCCs** and spectrogram features from **1,000+ audio samples** using Librosa.
- **Optimization**: Performed automated hyperparameter tuning using **Optuna**, evaluating **1,000+ parameter combinations** to maximize model performance.

Certifications

Machine Learning Specialization – DeepLearning.AI