Vaidik Yadav

+91 7974476349 | vaidiky90@gmail.com

 Vaidik Yadav | Vaidik26 | Portfolio

Jabalpur, Madhya Pradesh - 482001, India

OBJECTIVE

Data Scientist & AI Engineer with 1.5+ years of experience building **LLM, NLP, and ML** systems for document intelligence and automation. Experienced in **end-to-end model development, deployment, and optimization** for scalable GenAI products.

SKILLS

- **Programming Languages:** Python, SQL
- **Machine Learning & AI:** Scikit-learn, CatBoost, XGBoost, LightGBM, TensorFlow, Transformers, Hugging Face, LangChain, LlamaIndex, RAG, Fine-tuning, OpenAI APIs
- **Data Engineering:** Pandas, NumPy, Feature Engineering, Hypothesis Testing, ETL, SMOTE, PCA, Time Series
- **MLOps & Deployment:** Docker, DVC, MLflow, Flask, FastAPI, GitHub Actions, CI/CD, Model Monitoring
- **Cloud & Databases:** AWS (EC2, S3, Bedrock), GCP, MySQL, MongoDB, FAISS, Chroma, Pinecone
- **Visualization & BI:** Matplotlib, Seaborn, Plotly, Power BI, Excel

EXPERIENCE

- **Platinus Technology Pvt Limited** *September 2022 – August 2024*
  Data Analyst Hyderabad, India
  ○ Automated job-market data analysis using Python and SQL, reducing manual reporting by 60%.
  ○ Designed Power BI dashboards for client performance tracking, improving decision-making speed by 40%.
  ○ Reduced data retrieval latency by **30%** through SQL indexing and query optimization.
  ○ Collaborated across data and strategy teams to optimize performance metrics, achieving a **15% increase in overall efficiency**.

EDUCATION

Gyan Ganga College Of Technology
Bachelor of Technology (B.Tech) – Computer Science
Jabalpur, Madhya Pradesh
2021 – 2025

PROJECTS

Document QnA – LLM-powered Contextual Search GitHub Repository
Tools: LangChain, FAISS, Streamlit
- Built a **document-based QnA system** enabling context-aware retrieval from PDFs using RAG.
- Improved retrieval accuracy by **25%** using FAISS embeddings and hybrid search optimization.
- Deployed a **Streamlit** app for real-time user interaction with document queries.

Legal Advisory Chatbot – AI-powered Legal Assistant GitHub Repository
Tools: LlamaIndex, Hugging Face, Flask, RAG
- Engineered an LLM-powered chatbot to retrieve and summarize legal case files with **85% retrieval precision**.
- Optimized inference **latency under 2s for 100+ concurrent users** via Flask backend tuning.
- Integrated domain-specific contextual embeddings to enhance accuracy in contract interpretation tasks.

Milk Quality Classification – ML Pipeline Automation GitHub Repository
Tools: FastAPI, Scikit-learn, Docker, DVC, MongoDB
- Delivered **89% accuracy** using Gradient Boosting with **RandomizedSearchCV optimization**.
- Automated data preprocessing and feature encoding for scalability.
- Built a **version-controlled ML pipeline** with DVC and Docker for reproducible deployment.

CERTIFICATIONS & COURSES

- **Data Analytics – PW Skills**
- **Data Science Noob to ProMax – 5 Minutes Engineering**