

Vaidik Yadav | Data Scientist

+91 7974476349 | vaidiky90@gmail.com

 [Vaidik Yadav](#) |  [Vaidik26](#) | [Portfolio](#)

Jabalpur, Madhya Pradesh - 482001, India

OBJECTIVE

Data Scientist with 1.5 years of analytics experience designing and deploying ML, NLP, and GenAI solutions that drive measurable business outcomes. Proven track record in building end-to-end pipelines, optimizing models for speed and accuracy, and deploying production-ready applications with Docker, FastAPI, and CI/CD. Skilled in LLMs, RAG, and vector search for real-time contextual intelligence.

SKILLS

- **Programming Languages:** Python, SQL
- **ML/DL/NLP/GEN-AI:** Scikit-learn, CatBoost, XGBoost, LightGBM, Random Forest, Gradient Boosting, TensorFlow, LSTM, Transformers, Hugging Face, LangChain, OpenAI APIs, Supervised & Unsupervised Learning
- **Data Processing:** Pandas, NumPy, Feature Engineering, EDA, Hypothesis Testing, Time Series Analysis, SMOTE, A/B Testing, PCA, ETL
- **MLOps & Deployment:** Flask, FastAPI, Docker, DVC, GitHub Actions, MLflow, CI/CD, Model Deployment & Monitoring
- **Cloud Databases:** AWS (EC2, S3, SageMaker), MySQL, MongoDB, Pinecone, FAISS
- **Visualization BI:** Matplotlib, Seaborn, Plotly, Power BI, Excel

EXPERIENCE

- **Platinus Technology Pvt Limited** September 2022 – August 2024
Data Analyst Hyderabad, India
 - Increased client placement visibility by 40% through interactive Excel and Power BI dashboards, enabling faster decision-making.
 - Boosted placement rates by 20% by analyzing visa data and implementing trend-based targeting strategies.
 - Reduced job data retrieval time by 30% via database indexing and SQL query optimization.
 - Automated monthly reporting templates, saving 15+ hours of manual work per month for the analytics team.
 - Partnered with cross-functional teams to implement data strategies, leading to a 15% uplift in team performance.

PROJECTS

Medical Chatbot

 [GitHub Repository](#)

Tools: LangChain, LLMs, RAG, Pinecone, Flask, Hugging Face

- Delivered 90% accuracy on 200+ medical FAQs using RAG-based architecture
- Served 100+ concurrent requests with <1.5s latency through Flask REST APIs and prompt routing.
- Embedded 10,000+ medical text chunks for semantic search, improving contextual retrieval precision.

Milk Quality Classification

 [GitHub Repository](#)

Tools: Python, Scikit-learn, Docker, DVC, MongoDB

- Achieved 89% accuracy on multiclass classification using Gradient Boosting RandomizedSearchCV.
- Built reproducible ML pipeline with DVC and Docker for versioned deployment.
- Automated preprocessing with label encoding and missing value handling for scalability.

Stock Price Prediction (LSTM)

 [GitHub Repository](#)

Tools: TensorFlow, Pandas, Matplotlib

- Enhanced prediction stability by 18% via advanced time-series preprocessing & sliding windows.
- Prevented temporal leakage using walk-forward validation for robust model performance.
- Visualized predicted vs. actual prices for the last 100 trading days to validate accuracy.

EDUCATION

Gyan Ganga College Of Technology
Bachelor of Technology (B.Tech) – Computer Science
Jabalpur, Madhya Pradesh
01/2021 – 01/2025

CERTIFICATIONS & COURSES

- **PW Skills Data Analytics**
- **5 Minutes Engineering Data Science Noob To ProMax Batch 5**