SURAJ KUMAR

[redacted] | [redacted] | LinkedIn | GitHub

PROFESSIONAL SUMMARY

AI and Data Science enthusiast with practical experience in Data Science, Machine Learning, Deep Learning, Generative AI, and Agentic AI applications. Skilled in Python, SQL, statistical analysis, model development, and AI application deployment. Experienced in building end-to-end projects involving predictive analytics, LLM-powered systems, RAG architectures, AI agents, and MLOps practices. Passionate about transforming data into actionable insights and developing intelligent AI solutions that create real-world impact.

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, C, JavaScript
AI / LLMs: LangChain, LangGraph, Hugging Face, RAG, Gemini, Prompt Engineering
Data Science: Pandas, NumPy, Machine Learning, Deep Learning, NLP, Transformers, EDA, Feature Engineering, Statistical Analysis
Backend & Databases: FastAPI, Flask, Django, MySQL, MongoDB, PostgreSQL
Frontend: React.js, HTML, CSS, JavaScript, Bootstrap
MLOps & Tools: MLflow, DVC, Dagshub, Docker, Git, GitHub, Jupyter Notebook, VS Code
Core Concepts: Statistics, Probability, Model Evaluation, Data Preprocessing

EXPERIENCE

AI Engineer & Full-Stack Developer — Freelancing **May 2025 – May 2026**

- Built 3D scroll-based product websites using video frame pipelines (Whisk + Three.js) for 5+ client campaigns, cutting design turnaround by ~40%.
- Developed automated lead generation systems (Google Maps + Reddit scraping) with AI-powered personalised WhatsApp/email outreach, generating 200+ qualified leads per run.
- Architected GEW Intelligence OS — a multi-phase AI pipeline for inbound sales call analysis using Faster-Whisper transcription, Gemini reasoning, and FastAPI query layer.
- Delivered multi-tenant AI chatbot platforms and VAPI voice agents across gym, restaurant, and coaching verticals serving 3+ live clients.

Machine Learning Engineer — Suvidha Foundation | Remote **Feb 2025 – Apr 2025**

- Extracted and structured data from 50+ page research papers to build feature-rich ML training pipelines.
- Trained a LightGBM classifier achieving 91% accuracy on held-out test data; deployed as a FastAPI REST endpoint integrated with an internal dashboard, reducing manual reporting time by ~60%.

PROJECT EXPERIENCE

Autonomous Research Agent | *LangGraph · Qdrant · Hugging Face · GCP · FastAPI · React*

- Multi-agent platform where agents collaboratively plan queries, retrieve documents from 10K+ indexed Qdrant vector store, and synthesise answers with streamed output on GCP-deployed React frontend.

Fine-Tuned LLM for Exercise Coaching | *Llama 3 · PEFT · LoRA · QLoRA · PyTorch · Colab*

- Fine-tuned Llama 3 on biomechanics dataset using LoRA/QLoRA, reducing GPU memory by ~65%; achieved >85% form classification accuracy with personalised corrective exercise feedback generation.

MSME Operations Portal | *React · FastAPI · PostgreSQL · Google Gemini API · Scikit-Learn · Docker*

- Production-ready order management platform for eyewear operations, integrating a React analytics dashboard with a FastAPI/PostgreSQL backend to digitize fulfillment workflows.
- Integrated an AI-driven prescription parser using Google Gemini 3.5-flash to automatically extract structured clinical parameters (SPH, ADD power) from raw PDFs/images.
- Deployed a Random Forest classification model to predict SLA breach probability from real-time signals, triggering automated alerts for high-risk orders (>80% probability); containerized via Docker for one-command deployment.

Gym Helper | *OpenCV · MediaPipe · XGBoost · FastAPI · React · PostgreSQL*

- Real-time 30 FPS workout form evaluator using MediaPipe pose landmarks; XGBoost classifier at 88% accuracy with FastAPI microservices for inference and PostgreSQL-backed reporting.

GEW Intelligence OS | *Faster-Whisper · Gemini · FastAPI · Python · NLP*

- Multi-phase AI pipeline for inbound sales call analysis — transcribes Hinglish call recordings via Faster-Whisper, applies Gemini reasoning for voice intelligence, and serves insights through a FastAPI knowledge/query layer.

CredSafe — End-to-End Credit Card Fraud Detection | *Python · Scikit-learn · FastAPI · DVC · MLflow · Dagshub · GCP · React*

- Built an end-to-end MLOps pipeline for real-time fraud detection achieving ROC-AUC above 0.99; integrated DVC + Dagshub for versioning and MLflow for experiment tracking, deployed via FastAPI/React with CI/CD-ready GCP Storage integration.

CERTIFICATIONS

Agentic AI & Multi-Agent Systems — LinkedIn Learning	2026
LLM & RAG Pipeline Development — LinkedIn Learning	2026
Prompt Engineering for Generative AI — LinkedIn Learning	2025
Data Science & Machine Learning Fundamentals — LinkedIn Learning	2025
Python for Data Science and AI Development — LinkedIn Learning	2025

EDUCATION

BS — Data Science & Applications | IIT Madras Expected 2027