Surya Prakash Baid

linkedin.com/suryaprakashbaid github.com/surya-sgit suryaa.baid@gmail.com +91 9073332551

Data Scientist specializing in Classical ML, DL, and Generative AI with experience in developing end-to-end AI applications. Focused on building scalable, business-aligned models that transform raw data into measurable operational efficiency.

Personal Projects

LegalBuddy: LLM-powered legal chatbot for Indian laws

- Developed a prototype GenAI legal assistant that delivers guidance on legal challenges and corrective steps.
- Built a structured knowledge base from 1,000+ publicly available case records and IPC documents, enabling accurate semantic retrieval.
- Empowered English-speaking Indian users (10–20% of the population) to access legal guidance instantly—achieving up to 80% time savings over traditional methods like manual research or early-stage lawyer visits.
- Engineered with a robust tech stack (DeepSeek-R1, ChromaDB, LangChain) and deployed via FastAPI and Streamlit, delivering a seamless real-time query interface.

Vehicle Number Plate Detection

- Built an AI-powered number plate recognition system using YOLO for plate detection and PaddleOCR for text extraction, achieving 85% combined accuracy on real-world image and video inputs.
- Deployed via a Streamlit interface for seamless processing of recorded vehicle footage, with average recognition latency under 2 seconds per frame.
- Conceptualized for parking automation to replace manual entry/exit logging, aiming to reduce check-in time by up to 50% and cut queue delays during peak hours.

Work Experience

Data Analytics Intern | Connecting Dream Foundation, Remote *Oct 2024 - Nov 2024*

- Performed exploratory data analysis (EDA) leveraging Python and SQL on 4 health datasets to uncover 5 actionable insights and key patterns to support data-driven decisions.
- Developed 2 interactive PowerBI dashboards for key performance metrics of various health datasets.

Structural Analysis Research Intern | MIT, Manipal *Jan 2024 - May 2024*

- Analyzed seismic response of a G+10 building with varied soft-storey levels (ground to top floor) using ETABS and Python, optimizing fluid viscous dampers for structural stability.
- Achieved 48% reduction in story shear and 40% reduction in peak displacement through data-driven damper coefficient optimization.

Skills

Programming: Python, SQL
Libraries & Frameworks: NumPy, Pandas, TensorFlow, PyTorch, LangChain, HuggingFace, Scikit-learn, FastAPI, Streamlit.
Tools & Platforms: Git, Docker, AWS, Power BI, ChromaDB
ML & AI: Supervised/Unsupervised Learning, Deep Learning, Gen AI, NLP, Computer Vision, RAG, MCP
Soft Skills: Project Ownership, Problem Solving, Self-motivation, Creativity

Education

Manipal Institute of Technology (MIT), Manipal *2020 - 2024*
Bachelor of Technology in Civil Engineering *CGPA: 8.20/10*
Minor: Data Science

Achievements & Certifications

- Selected among top 3,000 students across India for a learning program conducted by Amazon scientists (Amazon ML summer school).
- AI/ML for Geodata Analysis - ISRO
- Google Data Analytics Professional Certificate
- Data Science: Foundations using R (Johns Hopkins University)