Aviraj Virape

Data Science Engineer (Intern)

Pune, Maharashtra | +91 99218 83944 | virapeaviraj@gmail.com

linkedin.com/in/avirajvirape | github.com/aviraj1805 | aviraj1805.github.io/portfolio | leetcode.com/u/aviraj_virape

SUMMARY

Final-year B.E. AI & Data Science student; ex-founder (COEP I2I Startup Program); research presented at INMEC-2026. Builds end-to-end ML in Python: EDA, scikit-learn models, evaluation and Streamlit deployment; 275+ LeetCode problems.

EXPERIENCE

Founder | Mahaguru AI | GitHub Jan 2024 - Nov 2025

- Founded an AI student-guidance platform; selected into COEP's I2I Startup Program for mentorship and incubation.
- Designed a multi-agent LLM architecture (FastAPI, React) and curated a custom student-mentor dialogue dataset.
- Collected feedback from ~150 student users on early prototypes; presented the underlying research at INMEC-2026.

Data Science Intern | Innovexis Feb 2026 - May 2026

- Built an ETL pipeline loading PhonePe Pulse JSON data (2018-2024) into a 9-table MySQL database for SQL analysis.
- Developed a multi-page Streamlit and Plotly dashboard answering 10 business questions via SQL queries on MySQL.

PROJECTS

Fix with Me - AI Repair Assistant | *Python*, *FastAPI*, *Gemini*, *Gemma*, *RAG*, *Supabase*, *scikit-learn* GitHub | Live

- Built a Bayesian repair-diagnosis engine (English/Hindi/Hinglish): 90.7% top-1 accuracy vs 81.3% for plain Gemini.
- Engineered hybrid RAG on Supabase (pgvector + PGroonga, Reciprocal Rank Fusion; 100% recall@5) with per-claim LLM verification: 92.5% correct answers and 0% unsupported claims vs 87.5% and 34% for plain Gemini.
- Trained a TF-IDF + logistic regression safety classifier on 3.8K examples, exported to a 0.68 MB pure-Python runtime (~0.2 ms/message): hazard recall 67.2% to 89.1%; safety-stop recall 100% vs 84.6% for plain Gemini, 0% false stops.
- Deployed FastAPI and Docker (Render) and a Next.js PWA (Cloudflare Workers); CI: 118 pytest tests and a safety gate.

Customer Churn Prediction | *Python*, *XGBoost*, *LightGBM*, *scikit-learn*, *Pandas*, *Seaborn* GitHub | Live

- Trained XGBoost and LightGBM on ~594K telecom records (21 features), 5-fold stratified CV: 0.916 mean ROC-AUC.
- EDA churn drivers: e-check payment (49%), month-to-month (42%), fiber-optic (41.5%), first-year customers (49%).

IAMARS - Drone Detection and Tracking | *Python*, *YOLOv8*, *ByteTrack*, *NumPy*, *OpenCV* GitHub | Live

- Led a team fine-tuning YOLOv8n on VisioDECT (leakage-aware split): 0.965 mAP@0.5 on 1,800 held-out test images.
- Evaluated tracking with MOTA/IDF1 on 91 test videos and intercept-point prediction in 3D simulation; runs at 88.5 FPS on an RTX 2050 GPU and 20 FPS on CPU. Found and fixed a label-conversion bug that had inflated training mAP.

SKILLS

Languages & Core: Python, SQL (MySQL), Data Structures and Algorithms, Maths for AI

Machine Learning: scikit-learn, Supervised & Unsupervised Learning, Feature Engineering, Model Evaluation

Data Analysis: Pandas, NumPy, Data Preprocessing, EDA, Matplotlib, Seaborn, Plotly

Generative AI: LLMs (Gemini, Gemma), RAG, Prompt Engineering, LLM Evaluation

Backend & Deployment: FastAPI, Pydantic, Docker, Streamlit, Render, Cloudflare Workers, GitHub Actions (CI/CD), pytest

Databases & Tools: MySQL, Supabase, Git, GitHub, Jupyter Notebook, Google Colab, Kaggle

PUBLICATION

"AI-Powered Reflective Mentorship" | INMEC-2026, Keystone School of Engineering, Pune Mar 2026

1st International Conference on Interdisciplinary Advances in Mechanical, Electronics and Computer Engineering

EDUCATION

B.E. Artificial Intelligence and Data Science | APCOER, Pune | CGPA 9.28 Expected May 2027

HSC (Science) | Yashodhara Junior College, Solapur | 78% May 2022