

Suraj S Nair

Computer Science (AI & ML) Undergraduate | Data Analytics, Applied Machine Learning & Backend Development
+91 8341511136 | surajnair038@gmail.com | github.com/SurajN4ir | linkedin.com/in/SurajN4ir | Hyderabad, India

PROFESSIONAL SUMMARY

Computer Science undergraduate (AI & ML) with hands-on experience across data analytics, backend API development, and applied machine learning. Skilled in Python, SQL, and FastAPI, with a track record of cleaning and analyzing large datasets, building scalable data pipelines, and translating statistical findings into stakeholder-ready insights. Experienced designing and deploying production-style systems — from vector-indexed retrieval architectures to REST APIs — that turn raw data into reliable, measurable outcomes.

EDUCATION

Marri Laxman Reddy Institute of Technology and Management, Hyderabad

Mar 2023 – Mar 2027

Bachelor of Technology, Computer Science and Engineering — AI & ML

CGPA: 8.18 / 10

EXPERIENCE

Data Analytics & Data Science Industry Training Program — AISPRY

Aug 2025 – Mar 2026

- Took ownership of cleaning, transforming, and analyzing 50,000+ records using Python (Pandas, NumPy), delivering a ~20% improvement in data quality through systematic validation and outlier handling.
- Dove deep into 5+ enterprise datasets, applying statistical methods — hypothesis testing and correlation analysis — to uncover actionable business patterns.
- Built trend and volatility visualizations covering 2+ years of pricing data for 10+ pharmaceutical companies (Matplotlib, Seaborn), delivering insights stakeholders could act on directly.
- Authored 15+ data-driven reports, translating statistical analysis into clear, actionable recommendations that drove measurable business decisions.

Skills Used: Python, Pandas, NumPy, Matplotlib, Seaborn, Statistical Analysis

PROJECTS

RepoMind | [GitHub](#)

Apr 2026 – Jul 2026

Developer

- Owned the end-to-end build of a full-stack retrieval platform that indexes codebases into a searchable knowledge base, simplifying how developers query unfamiliar repositories.
- Dove deep into repository structure to engineer a data pipeline that parses source code into logical units, generates vector embeddings (FastEmbed), and loads them into a Supabase/pgvector database for low-latency semantic retrieval.
- Designed and built REST APIs (FastAPI) to serve a multi-agent orchestration layer (LangGraph) coordinating Groq and Ollama inference, delivering grounded, source-referenced responses.

Tools: Python, FastAPI, Next.js, FastEmbed, Supabase, pgvector, LangGraph, Groq

SmartSort AI | [GitHub](#)

Oct 2025 – Dec 2025

Developer

- Owned the end-to-end build of a full-stack desktop application that classifies and organizes files using semantic content analysis instead of static extension-based rules.
- Built a data-processing pipeline — metadata extraction, embedding generation (FastEmbed), and vector storage (ChromaDB) — delivering fast, content-aware semantic search.
- Implemented duplicate-detection algorithms (content and perceptual hashing) plus storage analytics, delivering actionable cleanup recommendations through an interactive interface.

Tools: Python, FastAPI, SQLite, Git, ChromaDB, FastEmbed

TECHNICAL SKILLS

Programming Languages: Python, Java, C, JavaScript, SQL

Data & Analytics: Pandas, NumPy, Statistical Analysis, Hypothesis Testing, Data Visualization, Feature Engineering, Data Cleaning

Machine Learning & AI: Scikit-learn, TensorFlow, Model Evaluation, RAG, LangGraph, LLM APIs

Backend & Systems: FastAPI, REST API Design, Docker, Git, Postman

Databases: MySQL, PostgreSQL (Supabase/pgvector), SQLite, Firebase, MongoDB

CERTIFICATIONS

McKinsey Forward Certificate — McKinsey & Company | Google AI & ML Virtual Internship — EduSkills

Cisco NetAcad Riders Competition — Bronze Medal | Pearson MePro C2 — CEFR Level C2