SOPAN MUNDE

Maharashtra, India | +91-8767338447 | sopanmunde5@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

AI/ML Engineer specializing in LLM applications, Retrieval-Augmented Generation (RAG), and AI backend systems. Experienced in building intelligent applications using Python, FastAPI, LangChain, and vector databases. Developed multiple AI solutions including chatbots, NLP systems, and data-driven applications. Passionate about building scalable AI products and solving real-world problems with machine learning.

SKILLS

- **AI / LLM Technologies**: CrewAI, Pytorch, LangChain, LlamaIndex, LangGraph, LangSmith, RAG Architecture, Prompt Engineering, Vector Databases (Pinecone, FAISS), Model evaluation (RAGAS), API rate limiting, OpenAI API, Gemini, Llama, Hugging Face, Agentic AI, LLM fine-tuning (LoRA)
- **Machine Learning**: Scikit-learn, TensorFlow, NLP, TF-IDF, Data Preprocessing, Feature Engineering
- **Programming**: Python, SQL, FastAPI, Next.js, REST API, API integration, Streamlit
- **Data & Visualization**: NumPy, Pandas, Power BI, Excel, Matplotlib, Seaborn
- **Dev Tools**: Git, GitHub, AWS, Docker(basic), CI/CD(basic)
- **Database**: MySQL, MongoDB, PostgreSQL

PROJECTS

Trishul-AI(medical Chatbot) | GitHub
Mar 2026 - Present

AI chatbot providing general medical information using RAG architecture.
Tech: Python, FastAPI, LangChain, Pinecone, Next.js
•Built a RAG-based medical chatbot indexing 3,000+ medical documents using Pinecone embeddings.
•Achieved ~85% retrieval accuracy with 2s API response latency using FastAPI backend, supporting 20+ test users.
•Developed a Next.js interface with authentication for secure and user-friendly interaction.

Web Scraper RAG AI | GitHub
Jul 2025 - Sep 2025

Automated AI system for extracting web data and generating contextual answers.
Tech: Python, LangChain, Pinecone, Streamlit, OpenAI API
•Built an automated web scraping pipeline collecting 5k+ web pages, converting content into vector embeddings stored in Pinecone.
Implemented RAG-based query system achieving ~85% response relevance with average latency under 2 seconds.

Plagiarism Detection | GitHub
Dec 2024 - Dec 2024

AI-based plagiarism detection system for identifying duplicate or highly similar text using NLP techniques.
Tech: Python, Scikit-learn, NLP, Pandas, NumPy
•Built an ML-based plagiarism detection system using TF-IDF vectorization and cosine similarity to compare textual documents.
•Processed 1,000+ text documents, achieving ~90% similarity detection accuracy for duplicate and paraphrased content.

EXPERIENCE

AIML Engineer Trainee
Mar 2025 - May 2025

Prushal Technology Pvt. Ltd.
•Worked on Python-based ML implementations and data analysis pipelines
•Developed machine learning workflows including data preprocessing and model evaluation
•Collaborated on AI-based predictive modeling projects, applying ML algorithms to generate data-driven insights.

Data Science Intern
Feb 2025 - Sep 2025

Internship Studio
•Worked with Python, ML algorithms, and MySQL for data analysis
•Built classification and regression models using Python and Scikit-learn, improving predictive analysis on structured datasets.
•Automated data analysis workflows with Python, reducing manual data processing effort and improving analysis efficiency.

EDUCATION

PES College of Engineering
Chh. Sambhajinagar, Maharashtra, India

Bachelor of Technology, Computer Science and Engineering
Dec 2021 - Jul 2025
CGPA**:** 7.08

CERTIFICATIONS

•**MISAC, California-USA:** Certified Python Programmer Associate.
•**AWS Academy:** AWS Cloud foundation.
•**AWS Academy:** AWS Cloud Architecture.
•**Brainovision Solutions India Pvt. Ltd:** Data Science using Python.