Soham Deshpande

+91-9307005165 | sohamdeshpande242@gmail.com | LinkedIn | GitHub

Education

Indian Institute Of Information Technology, Nagpur, Maharashtra, India
B.Tech in Electronics and Communication Engineering, CGPA: 8.24   Dec 2021 – May 2025

Achievements

Competitive Programming: LeetCode 1694 rating (**Top 13.24%**), CodeChef **3-star** (1794 max rating); **Global Rank 17** in CodeChef Starters 108 (21k+ participants); Solved **750+ DSA/CP problems.**

Hackathons and Open Source: Winner at **Version-Beta Hackathon, MANIT Bhopal**; Regional Winner, **Smart India Hackathon (SIH)**; Contributed BVPS and GST functions to a fintech API in GSSOC.

Experience

AI/ML Intern, SPR & CO   July 2024 – Present

- Built a **Contract Type Detection** and **Clause Type Detection Model** using **LegalBERT**, achieving a **96.8% accuracy rate**, improving contract classification efficiency by **20%**.
- Engineered a **RAG** (Retrieval-Augmented Generation) approach integrated with **Pinecone DB**, enhancing clause span identification and type detection accuracy by **15%**.
- Deployed **model tracking and versioning** with **MLflow** and integrated models into a **Next.js app** via **FastAPI**, resulting in a **30%** reduction in integration time.
- Currently advancing **contract risk assessment** using a **Multi-AI Agentic** approach with fine-tuned domain-specific models.

Machine Learning Intern, Drooid   March 2024 – May 2024

- Processed a **210K-headline dataset**, extracting features with **TF-IDF** and **Word2Vec**, leading to a **25%** growth in classification accuracy for news categorization.
- Fine-tuned **DistilBERT** for news classification, achieving high accuracy, visualized category distributions with word clouds, and optimized classification workflows.

Projects

News AI Agent   GitHub Link

- Built a **News AI Agent** using **CrewAI** and **Google Gemini LLM**, integrating tools like Serper and Wiki for real-time news generation.
- Created agents like **Researcher** and **Writer** to deliver accurate and **97% relevant news summaries**.
- Enhanced response speed and accuracy by **20%** through system optimizations and workflow tuning.

Facial Emotion Detection with CNN   GitHub Link

- Developed a facial emotion detection system using CNNs, achieving **95%** accuracy with ResNet50 architecture.
- Enhanced model robustness through normalization, ReLU, and data augmentation techniques.
- Utilized transfer learning with **VGG16 and ResNet50** to boost performance.

Cold-Mail-Generator   GitHub Link

- Created an end-to-end LLM-based cold email generation tool to help Software and AI services companies automate outreach.
- Integrated **Llama 3.1 LLM**, **ChromaDB**, **LangChain**, and **Streamlit** to provide a seamless user experience for generating personalized cold emails.
- Built a system to input job portal links and generate tailored emails for HR, improving outreach efficiency.

Other Projects   GitHub Link

- **Text-to-SQL App, Finetuned LLAMA2, ChatwithPDF, Course Recommendation System**

Technical Skills

- **Languages:** C/C++, Python, JavaScript, SQL, NoSQL, Git, Matlab, Linux, Bash
- **Tools and Frameworks:** Pandas, Numpy, Scikit-learn, Keras, PyTorch, TensorFlow, OpenCV, LangChain, Django, Flask, AWS, GCP, Tableau, GitHub
- **Coursework:** DSA, OOP, Machine Learning, OS, DBMS, Deep Learning, Computer Vision, NLP