Sandeep Chaurasiya

Bangalore, Karnataka | [redacted] | [redacted] | LinkedIn | GitHub

Profile Summary

Data Science & AI student at IIT Guwahati with strong foundations in Machine Learning, Deep Learning, NLP, and Generative AI. Experienced in building end-to-end AI systems including RAG pipelines using LangChain and Hugging Face Transformers. Skilled in data preprocessing, model evaluation, and deploying scalable LLM-powered solutions using FastAPI and Streamlit. Passionate about applying AI to solve real-world problems.

Education

Indian Institute of Technology Guwahati — Guwahati, Assam
B.S. in Data Science & Artificial Intelligence | CPI: 8.66/10.0 — *Sep 2023 – Present*

Rani Laxmi Bai Memorial School — Lucknow, Uttar Pradesh
Class X: 96.8% | Class XII: 94.8% — *Sep 2019 – March 2023*

Technical Skills

Languages: Python, SQL
Machine Learning: Regression, Classification, Decision Trees, Random Forest, XGBoost
Deep Learning: ANN, CNN, Transfer Learning, Early Stopping
NLP: Tokenization, Text Preprocessing, Embeddings, Transformers
Libraries: NumPy, Pandas, Scikit-learn, TensorFlow, PyTorch, Hugging Face Transformers, LangChain
Generative AI: RAG, LLM Fundamentals, Prompt Engineering
Tools: FastAPI, Streamlit, Docker, Jupyter Notebook, VS Code, Power BI

Projects

Healthcare Premium Prediction | *Python, Scikit-learn, XGBoost, Streamlit* | GitHub

- Built an end-to-end ML system to predict health insurance premiums using demographic and medical data.
- Engineered features and reduced multicollinearity using VIF, trained XGBoost, Linear, and Ridge models achieving $R^2 \approx 0.98$.
- Designed age-segmented models (≤ 25 and > 25) and deployed a **Streamlit** app for real-time predictions.

Medical RAG Chatbot | *LangChain, FAISS, LLaMA 3.1, Hugging Face, Streamlit* | GitHub

- Built a Retrieval-Augmented Generation (RAG) system using **LangChain, FAISS, and LLaMA 3.1** for accurate medical Q&A.
- Used **Hugging Face sentence-transformers** for semantic embeddings and FAISS for efficient vector similarity search.
- Reduced hallucination by grounding responses strictly on retrieved medical context; deployed on **Streamlit Cloud** with secure API handling.

Brain Tumor Classification using CNN | *TensorFlow, OpenCV, NumPy, Streamlit* | GitHub

- Developed a CNN model to classify 7000+ MRI images into four tumor categories with **96% validation accuracy**.
- Applied preprocessing techniques: normalization, label encoding, and one-hot encoding; used Early Stopping to prevent overfitting.
- Architected and deployed a real-time MRI classification web app using **Streamlit**.

Certifications

- **Oracle Cloud Infrastructure 2025** – AI Foundations Associate
- **SQL Basics to Advanced** – SkillNation

Achievements

- **Agentic AI Hackathon** – Participated among 100+ teams at IISc Bangalore
- **2-Star Coder on CodeChef** – Data Structures and Problem Solving
- Solved **300+ DSA problems** across competitive programming platforms