SOURAV KUMAR

souravkumar8432@gmail.com +916205208204 Hyderabad, Telangana, India

iamsourav.xyz github.com/singhsourav0 linkedin.com/in/singhsourav0

ABOUT

Python developer with hands-on experience in Data Science, ML pipelines, and Deep Learning. Built and deployed data-driven applications using Flask and Streamlit. Designed end-to-end ML workflows with Docker, DVC, MLflow, and AWS EC2. Worked on scalable data processing using Spark and Databricks. Integrated GitHub Actions for CI/CD automation. Developed LLM-based apps using LangChain, Hugging Face, and FastAPI, incorporating RAG pipelines and vector search with ChromaDB.

EXPERIENCE

Data Analytics Job Simulation, *Forage* June – July 2025

- Completed a Tata iQ job simulation on forage that focused on AI driven data analysis, leveraging GenAI for EDA, risk assessment and strategic insights.
- Proposed a predictive model for delinquency risk, designed an AI-powered collections strategy, and created a stakeholder focused report.

Virtual Internship, *Future Skills Prime* 03/2023 – 06/2023

- During the Microsoft Future Ready Talent virtual internship, I built a Face Mask Detection system using YOLOv5 for object detection and integrated real-time video processing with OpenCV achieving 92% accuracy. Deployed the end-to-end ML pipeline on Microsoft Azure.

PROJECTS

AutoStocks : AI-Powered Stock Market Report Generator, *https://autostocks.onrender.com/*

- Built a Flask app to automate stock analysis from user-submitted stocks and deliver PDF reports via email.
- Built data pipelines to collect and analyze data for all users from multiple sources, creating detailed charts and reports. Integrated Gemini AI for sector wise insights, along with NLP for news summarization.
- Automated the entire workflow using GitHub Actions to generate and schedule emails with PDF reports for all users. it reduces manual effort by 90% for retail investors.

Real Estate Analytics, *https://houseing.onrender.com/*

- Built a Flask-based ML website to solve real-world property challenges in the Gurugram housing market.
- Scraped data from 99acres, built a price predictor (R^2: 0.90, MAE: 0.3) with 12k cross-validation runs across GBM and RF, and integrated hybrid recommendation systems for personalized insights.
- Built an interactive analytical dashboard with geolocation heatmaps, sector wise charts, price trends and more insights to analyze real estate patterns in Gurugram.

Chat with Website : RAG-based Assistant, *https://chatwebsite.streamlit.app/*

- Built a real-time AI assistant using BeautifulSoup, Gemini API, and ChromaDB with a RAG pipeline to extract website content, generate embeddings, and perform semantic search for context-aware answers.
- Deployed an interactive Streamlit app that offers a user-friendly chat interface, enabling real-time Q&A from any website URL.

Kidney Disease Classification, *https://github.com/singhsourav0/Kidney-Disease*

- Built a CNN model to classify kidney disease from medical images, achieving 95% accuracy by optimizing parameters. Implemented MLOps using MLflow for experiment tracking, DVC for data versioning, and CI/CD pipelines for seamless automation.
- Containerized the application with Docker and deployed it on AWS EC2, enabling scalable, real-time access to the classification model.

EDUCATION

B.Tech in Computer Science Engineering, 2021 – 2025
Bihar Engineering University

SKILLS

Programming Languages — Python, C++ , SQL

Data Science & Analysis — Pandas, NumPy, Matplotlib, Seaborn, Pyspark, Scikit-learn, XGBoost, TensorFlow, Keras, NLTK, LangChain

MLOps & Deployment — Docker, GitHub Actions, MLflow, DVC, AWS EC2, Firebase, databricks

Web & Dev Tools — Flask, FastAPI, HTML, CSS, Streamlit, Git, GitHub, Selenium, Requests

CERTIFICATION

NPTEL Online Certification — Data Mining and Deep Learning

COURSERA — IBM Data Science Specialization