

SOUKARYA RAHA

Raiganj, West Bengal ☎ +91-8972383015 ✉ soukaryaraha@gmail.com 🌐 LinkedIn 🐙 GitHub

SKILLS

Languages	: Python, SQL
Libraries	: Pandas, NumPy, SciPy, Matplotlib, Seaborn, TensorFlow, Keras, ScikitLearn
Frameworks	: LangChain, Streamlit, Flask, FastAPI
AI & ML	: Machine Learning, Deep Learning, Computer Vision, NLP, LLMs, Generative AI
Database Management	: SQL (SQL Server, MySQL), MongoDB
Data Visualization	: Tableau, Power BI, Excel
Tools & Platforms	: MLflow, Postman, Jupyter Notebook, VS Code, Hugging Face, Tender

EDUCATION

Vellore Institute of Technology	Bhopal, India
Master of Technology in Data Science and Artificial Intelligence	Aug 2023 - Expected Sep 2025
CGPA: 8.63/10	
Sapthagiri College for Engineering	Bengaluru, India
Bachelor of Engineering in Biotechnology	Jul 2019 - Jun 2023
CGPA: 9.2/10	

WORK EXPERIENCE

Data Analyst Intern	Sep 2024 – Oct 2024
NullClass Edtech Pvt Ltd 🔗	Bhopal, India
<ul style="list-style-type: none">Cleaned and preprocessed job portal data using NumPy and Pandas, achieving 95% data accuracy.Optimized Tableau filters and analyzed 10,000 job listings to identify trends that boosted user engagement by 30%.Developed interactive geographic maps for 60+ job listings, integrating click-to-zoom and custom color coding to enhance regional insights and job matching algorithms by 20%.Collaborated with cross-functional teams to design and implement data-driven solutions leading to a 25% increase in job application submissions.	

PROJECTS

Liver Disease Prediction: A Comparative ML Analysis 🔗 <i>MLflow, Flask, Docker</i>	Oct 2024
<ul style="list-style-type: none">Processed 584 records with 15 features, applying hypothesis testing, feature engineering, SMOTE for imbalance handling, and EDA for distribution and correlation analysis.Optimized 5 ML models using Optuna, improving accuracy by 30%. SVM demonstrated the best performance in early disease detection. Logged 6 key metrics in MLflow tracking confusion matrix, AUC-ROC, precision and recall.Built a Flask app with 10 inputs, containerized with Docker, automated via GitHub Actions, and hosted on Render with API tested in Postman.	
Emotion Recognition Using CNN 🔗 <i>CNN, FastAPI, Streamlit, Render</i>	Sept 2024
<ul style="list-style-type: none">Preprocessed 300 images using resizing, normalization, and augmentation with enhanced model robustness and reduced over fitting by 15%.Designed and fine-tuned a CNN model, achieving higher accuracy with lower loss, with 100% precision and recall, improving real-time emotion detection speed by 30%.Deployed a robust FastAPI application on Render using GitHub Actions, enabling smooth image uploads while maintaining response times under 2 seconds through the implementation of advanced optimization techniques that enhanced performance.	
Fine-tuning LLaMA 2 with QLoRA for text generation 🔗 <i>LLaMA 2, Hugging Face, TensorBoard</i>	Dec 2023
<ul style="list-style-type: none">Loaded the LLAMA-2-7B-CHAT model from Hugging Face and fine-tuned on MLABONNE/GUANACO-LLAMA2-1K of 1,000 samples using QLoRA with rank 64 and a scaling parameter of 16.Applied 4-bit precision (NF4) quantization and Parameter-Efficient Fine-Tuning (PEFT) to minimize VRAM usage while maintaining model performance. Trained the model for one epoch, tracking loss convergence and performance metrics via TensorBoard plots.Implemented a text generation pipeline, formatting inputs per LLAMA 2 prompt templates to ensure coherent and context aware responses.	

CERTIFICATIONS

- Data Analysis Intern - *Cognifyz Technologies*
- Career Essentials in Generative AI - *Microsoft & LinkedIn*
- Data Analyst Training - *NullClass EdTech Pvt. Ltd*
- Scientific Computing with Python - *Free Code Camp*
- SQL Skills Certification – *Hacker Rank*