Sourish Chatterjee

Location: Kolkata, WB, India
My Portfolio | LinkedIn | GitHub | Email: contact.sourishchatterjee@gmail.com | Mobile: +91 93303 19122

Career Objective

Aspiring Data Scientist with expertise in data preprocessing, feature engineering, and predictive modeling. Passionate about utilizing data-driven insights and machine learning techniques to solve complex problems and drive innovation.

Technical Skills

Languages	:	Python, Java, C
Tools & Libraries	:	Pandas, Numpy, Matplotlib, Scikit-learn, Streamlit
Data Analysis	:	Data Preprocessing, Data Cleaning, Data Visualization
Platforms	:	Google Cloud, GitHub
Tools	:	Visual Studio Code, Google Colab, Jupyter Notebook, Git (basic)

Projects

Nifty 50 Stock Prediction Model *Python, Numpy, Pandas, Scikit-Learn, VS Code, Streamlit* Source Code
Live Model

- Built and deployed a live stock price prediction model using Streamlit.
- Utilized Random Forest for predictive analysis, achieving over 90% accuracy on test data.
- Enhanced model performance through advanced feature engineering and optimization techniques.

Real Estate Price Prediction Model *Google Colab, Python, Numpy, Pandas, Matplotlib, Scikit-Learn* Source Code

- Preprocessed data using One-Hot Encoding, improving model accuracy by 15%.
- Trained a Random Forest Regressor achieving an R^2 score of 85% and Mean Squared Error (MSE) of 150,000.
- Optimized hyperparameters and enhanced feature selection, reducing prediction error by 10%.

Oil Spill Detection Using AIS Data and SAR Images (Hackathon Project) *Working On...* Source Code

- Focused on building an AIS-based anomaly detection model for maritime oil spills.
- Developed and tested ATS models, specifically Autoencoders and currently working on LSTM architectures, to identify anomalous ship movements.
- Contributed to the internal round qualification for the Smart India Hackathon 2024

GPT from Scratch *Working on..* Source Code

- Built a custom GPT model inspired by the "Attention Is All You Need" research paper.
- Trained the model on the **tinyshakespeare dataset**, generating text character by character.
- Designed and implemented encoding/decoding, dataset preparation, and tokenization using PyTorch.
- Explored foundational Bigram models and self-attention mechanisms, incorporating learnings from Andrej Karpathy's "Zero to Hero" series.

Certifications

Google Cloud Platform || Generative AI || MOOCs

Education

Meghnad Saha Institute of Technology Kolkata, WB, India
Bachelor of Technology in Computer Science Engineering with specialization in AIML *2022 – 2026*

Delhi Public School Ruby Park Kolkata, WB, India
Higher Education in Computer Science *2020 – 2022*