Swapnil Roy

Phone: +91 95315 34066
Email: swapnil.roy@msds.christuniversity.in
swapnilroydata@gmail.com
Address: Nagasandra, Bengaluru, India, 560073
LinkedIn | GitHub | Kaggle

Summary:

Data Science Master's student specializing in end-to-end AI/ML solutions across healthcare, finance, and enterprise analytics. Proven track record developing production-ready systems including medical RAG chatbots, computer vision for brain tumor detection (91.1% accuracy), and real-time fraud detection with PySpark. Expert in MLOps, explainable AI (SHAP), and full-stack deployment using Docker, FastAPI, and modern LLM technologies. Strong ability to translate complex business problems into scalable AI solutions with measurable impact.

Education:

- **CHRIST (Deemed to be University)**
  M.Sc in Data Science
  Ongoing

- **Bidhannagar College (WBSU)**
  B.Sc in Statistics
  CGPA: 8.7

- **St. Paul's Academy, Burdwan**
  ISC Board
  Class 12 Percentage: 80.5

- **St. Paul's Academy, Burdwan**
  ICSE Board
  Class 10 Percentage: 84.4

Experience:

- **Research Intern**
  Techno International New Town
  Deployed an HR system that predicts candidate hiring with 93% accuracy and uses XAI (SHAP) and Counterfactual (DiCE) to give hiring suggestions when rejected.

- **Data Science Intern**
  CodeClause Internship
  Developed and implemented data analysis techniques, achieving high accuracy in predictive models. Enhanced decision-making through data cleaning, visualization, and statistical modeling

Project:

- **MedQuest: End-to-End Medical Research RAG Chatbot**
  Github: https://github.com/S-T-R-A-N-G-E-R/MedQuest
  Built full-stack RAG application with Llama 3, FAISS vector store, and Multi-Query Retriever. Processes PDF/DOCX/CSV files via privacy-focused Streamlit interface..

- **Credit Card Fraud Detection System using PySpark**
  Github: https://github.com/S-T-R-A-N-G-E-R/Credit_Card_Fraud_Detection_with_PySpark
  Built fraud detection system using Gradient Boosted Trees on large-scale data. Deployed real-time FastAPI with monitoring dashboards, achieving 94% accuracy

- **Candidate Hiring Predictor**
  Github: https://github.com/S-T-R-A-N-G-E-R/hiring_pred_shap_CF
  Built HR analytics system using Random Forest with 93.3% accuracy, integrated SHAP for explainability and DiCE for counterfactual suggestions. Deployed via Streamlit.

- **Customer Churn Prediction System**
  Github: https://github.com/S-T-R-A-N-G-E-R/churn_prediction_app
  Designed end-to-end ML application with 98.3% accuracy, deployed on FastAPI (backend) and React (frontend), including real-time explainable AI insights (SHAP).

- **Brain Tumor Detection System using YOLOv8**
  Github: https://github.com/S-T-R-A-N-G-E-R/Brain_Tumor_Detection
  Developed an AI-powered medical imaging solution using YOLOv8 nano to detect and classify three types of brain tumors (Glioma, Meningioma, Pituitary) from MRI scans. Achieved 91.1% mAP50 accuracy with 11.7ms inference time. Implemented complete ML pipeline including data preprocessing, model training, validation, and deployment with comprehensive performance analytics.

Skills:

- **Programming Languages:** Python, R, JavaScript, SQL, HTML/CSS
- **Machine Learning:** Supervised/Unsupervised Learning, Computer Vision, NLP, Deep Learning, Feature Engineering, Model Evaluation
- ML Frameworks: PyTorch, TensorFlow, Keras, YOLOv8, Scikit-learn
- **AI/LLM Technologies:** RAG Systems, Vector Databases (FAISS), LLM Integration (Ollama, Llama 3), Explainable AI (SHAP, DiCE)
- **Big Data & Processing:** PySpark, Pandas, NumPy
- MLOps & Deployment: Docker, FastAPI, Streamlit, REST APIs, Model Monitoring, CI/CD
- **Data Engineering:** Database Design, MongoDB, Data Pipelines, ETL
- **Visualization & Monitoring:** Matplotlib, Seaborn, Prometheus, Grafana
- **Development Tools:** Git/GitHub, Jupyter, React (Frontend)
- **Statistical Analysis:** Hypothesis Testing, Regression Analysis, Time Series, A/B Testing

Trainings and Certifications:

- **Architecting Cloud Computing Solutions on Microsoft Azure**
  Infosys Springboard
  Basics on how Cloud Computing works, Hands on experience with Microsoft Azure
- **Generative AI with GPT**
  Infosys Springboard
  Live demonstrations of GPT applications like Auto HTML generation and IntelliSense.
- **OpenAI Generative Pre-trained Transformer 3 (GPT-3) for Developers**
  Infosys Springboard
  Explored GPT-3 architecture, practical NLP solutions, and Python integration.
- **AI-first Software Engineering**
  Infosys Springboard
  Integrated AI models into SDLC phases with an AI first approach.
- **Prompt Engineering**
  Infosys Springboard
  Learned prompt-based learning and GPT-3 task applications.
- **Python 101 for Data Science**
  Cognitive Classes, Powered By IBM Developer Skills Network
  Certificate ID: db4cadd1c19649b79a67c0ad0c7709b1
- **Programming with Python**
  UpGrad
  Certificate ID: 33048933

Positions of Responsibility:

- **IEEE**
  Chapter Vice-Chair
  Engage in professional development, networking, and knowledge-sharing within technology community

- **Tech ReVamp**
  Research Coordinator
  Working with teams and coordinating with other teams to lead new and innovative research in field of technology

- **International Model United Nation (IMUN)**
  Delegate
  Represented Pakistan for "Mental Health Action Plan: Promoting Care and Treatment"