SAMIR SENGUPTA

Data Scientist

New York, USA | +15513591228 | samir843301003@gmail.com | github.com/SamirSengupta | linkedin.com/in/samirsengupta/ | neuralthread.cloud/samir

SUMMARY

Data Scientist with a strong foundation in Deep Learning, Machine Learning, and Artificial Intelligence. Experienced in developing and deploying scalable AI solutions, optimizing machine learning models, and integrating cutting-edge technologies like LLaMA for code generation and security analysis. Proficient in computer vision, image manipulation, and deep learning algorithms. Proven track record in improving business outcomes through data-driven decision-making, statistical analysis, data modelling and model development. Skilled in data collection, data infrastructure, and the fine-tuning of models on unstructured data. Proficient in Python, TensorFlow, PyTorch, SQL, and Power BI, with a commitment to continuous learning and professional growth. Currently pursuing a Master's in Data Science to further enhance expertise in advanced data-driven methodologies.

EUCATION

MASTERS IN DATA SCIENCE. Sept 2024

Saint Peter's University. New Jersey, USA

Acquired Skills: Deep Learning, Machine Learning, Artificial Intelligence, Databases, Neural Networks, Large Language Models, Computer Vision, Data Modeling, Statistical Analysis.

BACHELORS IN DATA SCIENCE. April 2023

University of Mumbai. Mumbai, India

Acquired Skills: Python, SQL, Power BI, Tableau, Machine Learning, Deep Learning, Statistics, Predictive Models, Data Mining, Business Intelligence.

WORK EXPERIENCE

Research Assistant, Saint Peter's University. Nov 2024 – Present

- Leading research on open-source Large Language Models, leveraging Low-Rank Adaptation to enhance performance and adaptability.
- Building advanced tools with RAG (Retrieval-Augmented Generation) to enhance contextual relevance and efficiency.
- Developing tailored LLM solutions to support Saint Peter's University's unique academic and institutional objectives.

Software Developer, Synradar. Aug 2023 – July 2024

- Developed and optimized machine learning-based intrusion detection systems, improving threat detection accuracy by 40% and reducing response times by 25%.
- Automated and enhanced threat analysis processes using machine learning algorithms, increasing efficiency by 50%.
- Integrated LLaMA 3.1 for code generation and security evaluation, resulting in a 30% improvement in code quality and assessment efficiency.
- Designed and implemented interactive security analytics dashboards, leading to a 35% improvement in incident response times.

Data Scientist, Neural Thread. Jan 2022 – July 2023

- Developed and deployed machine learning models to predict key business metrics, increasing forecast accuracy by 25% and driving data-driven decision-making.
- Enhanced predictive model performance by 30% through advanced neural network architectures and hyperparameter tuning.
- Created generative AI models, such as chatbots and content generation tools, increasing customer engagement by 35%.
- Optimized AI solutions for production, reducing error rates by 40% and computational time by 50%, while delivering significant business value.

PROJECTS

Music Mate: Song Downloading System. Dec 2023

- Created a music downloader utilizing the Spotify API and Python Tube library to extract songs from Spotify playlists and YouTube videos.
- Constructed a Flask backend to manage API requests, enabling the downloading of songs based on user input.
- Developed a front-end interface for user interaction and input of Spotify playlist or YouTube video URLs.

Power BI: Sales Forecasting Dashboard. Jan 2024

- Created an interactive Power BI dashboard for sales forecasting, using advanced data analysis and visualization techniques to provide useful insights.
- Implemented powerful forecasting models to predict sales trends accurately, helping businesses optimize their strategies.
- Improved decision-making by presenting detailed sales analytics in an easy-to-use interface, making it simple for stakeholders to understand important insights.

Resume Evaluator: Gemini LLM based Candidate Shortlisting. March 2024

- Developed a Flask application utilizing Google's Gemini Large Language Model (LLM) to effectively summarize documents like CVs and job descriptions.
- Created robust functionality allowing the generation of concise summaries for both job descriptions and CVs, providing valuable insights into candidate suitability for hiring decisions.
- Integrated feedback generation features to recommend enhancements for candidates' professional profiles, thereby facilitating their career advancement.

MedScan.ai: Medical Image Recognition. Aug 2024

- Scans medical images to provide accurate and timely diagnoses, enhancing clinical decision-making.
- Combines visual and textual data inputs, offering comprehensive medical insights and a holistic view of patient health.
- Designed for seamless operation, making it easy for healthcare professionals to integrate into their workflows and improve patient care.
- Applied computer vision and pattern recognition techniques for accurate image analysis.

Jake.ai: Conversational AI. Nov 2024

- Jake.AI leverages advanced open-source large language models LLMs like LLaMA 3, Gemma 2, and Mistral for enhanced conversational and generative capabilities.
- It is locally hosted on your server using LM Studio, ensuring that all data remains private and secure.
- The platform functions as an AI companion, offering robust conversational abilities and innovative generative features.
- As an open-source solution, Jake.AI promotes transparency and flexibility in its AI interactions and implementations.

SKILLS &CERTIFICATIONS

- **Technical Skills**: Python (NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch), SQL (MySQL, PostgreSQL), ETL, Tableau, Power BI, Machine Learning, Deep Learning, Neural Networks, Data Processing, Artificial Intelligence, LLM (Large Language Models), Big Data, Generative AI, DevOps, Restful APIs, LangChain, RAG (Retrieval-Augmented Generation), API, Data Analysis tools, Data Management, Cybersecurity, Natural Language Processing, Predictive Analysis, Feature Engineering, Model Deployment, Reinforcement Learning, Multi-Modal Models, Computer Vision, R Programming, Data Engineering, Performance Optimization, Image Recognition, Data Infrastructure, Data Collection, Data Modeling, Statistical Analysis, Fine-tuning Models, Product Development, Data Visualization, Transformers, Business Intelligence, Pattern Recognition, Image Manipulation, Quantitative Research, BERT, KPIs.

- **Soft Skills**: Strong collaborative teamwork, effective communication, adept problem-solving, meticulous attention to detail, excellent time management, adaptable to change, committed to continuous learning, skilled in critical thinking, capable leadership, and analytical thinking, Decision Making.

- **Certifications**: Machine Learning from Stanford University, Generative AI from Google, Power BI, MySQL, Machine Learning, Large Language Models (LLMs) from DeepLearning.AI, Advanced Excel (Microsoft Office), Python, Deep Learning.