SAMIR SENGUPTA

Data Scientist

Mumbai, India | 8356075699 | samir843301003@gmail.com | github.com/SamirSengupta | linkedin.com/in/samirsengupta/ | neuralthread.cloud/samir

SUMMARY

Experienced Data Scientist proficient in Large Language Models (LLM), Machine Learning (ML), Artificial Intelligence (AI), and Deep Learning techniques. Skilled in Python libraries such as NumPy, Pandas, Scikit-learn, TensorFlow, and Keras, with expertise in advanced analytics and model development. Possesses a strong foundation in SQL databases, Tableau, and Power BI for effective data visualization and analysis. Capable of implementing cutting-edge algorithms and models to efficiently address complex challenges in the data science industry. Dedicated to leveraging AI, LLM, Lang Chain and data visualization skills to drive Innovation and continuous improvement in solving real-world problems.

EUCATION

BACHELORS IN DATA SCIENCE.   April 2023

University of Mumbai.   Mumbai, India

Acquired Skills: Deep Learning, Machine Learning, Artificial Intelligence, Databases, Neural Networks, Large Language Model.

Higher Secondary Education (Science).   April 2020

University of Mumbai.   Mumbai, India

Acquired Skills: Chemistry, Information Technology, Physics, Mathematics, Statistics.

WORK EXPERIENCE

Software Developer, Synradar.   Aug 2023 – July 2024

- Utilize Python and machine learning libraries, including TensorFlow, Keras, and PyTorch, to develop advanced neural network-based cybersecurity solutions.

- Integrate cutting-edge technologies and data visualization tools like Matplotlib and Seaborn to enhance threat detection and mitigation.

- Employ neural networks and machine learning techniques, leveraging packages such as Scikit-learn and NLTK, to proactively anticipate and address emerging cybersecurity threats.

- Implement AI technologies, including large language models like Meta's LLaMA 3 and Google's Gemini Pro, to automate and ensure comprehensive cybersecurity measures for esteemed clients.

Data Scientist, Neural Thread   Jan 2022 – July 2023

- Engineered ML models for precise customer churn prediction, resulting in a significant increase in customer retention rates from 71% to 93% annually.

- Applied cutting-edge deep learning algorithms to enhance the precision of Predictive Analytics Solutions, achieving a better prediction accuracy of 42%. Led K-means clustering initiatives for dynamic customer segmentation, optimizing targeted marketing strategies.

- Developed and implemented tailored retention models, resulting in a substantial increase in customer loyalty from 47% to 83%.

PROJECTS

Music Mate: Song Downloading System.   May 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 using HTML, CSS, and JavaScript 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.   Feb 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 Documents Scanner.   March 2024

- Developed an image recognition app using Google's Gemini Pro Vision model to analyze user-uploaded images.
- Incorporated advanced features to interpret images and provide descriptive prompts, assisting users in creating similar imagery.
- Integrated the Gemini Provision model to accurately discern image content, ensuring precise responses.
- Created an interactive user interface for easy image input and result display, enhancing user experience and engagement.

Jake.ai: Conversational AI.   July 2024
- Utilizes state-of-the-art open-source language models, including LLaMA 3, Gemma2, and Mistral.
- Enables natural, fluid conversations through advanced speech recognition and response capabilities.
- Customizable to specific needs and preferences for personalized interactions.
- Built on a robust foundation of open-source LLMs, leveraging community-driven advancements.
- Operates using local LLMs with LM Studio for enhanced privacy and control.

SKILLS &CERTIFICATIONS

- Technical Skills: Proficient in Python (NumPy, Pandas, Scikit-learn, TensorFlow, Keras), SQL (MySQL, PostgreSQL), Tableau, Power BI, R programming, with expertise in Machine Learning, Deep Learning, Neural Networks, Data Processing, Artificial Intelligence, LLM (Large Language Models), Generative AI, and Langchain, RAG (Retrieval-Augmented Generation).

- 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.

- Certifications: Google Certified Data Analyst, Generative AI from Google, Power BI, MySQL, Machine Learning, Large Language Models (LLMs) from DeepLearning.AI, Adobe Photoshop, Advanced Excel, Python, Machine Learning from Stanford University.