Saimanoj Bera

Saimanoj2509@gmail.com |+1 3134782240 | https://www.linkedin.com/in/saimanoj-bera-044831150/ | https://saimanojbera.github.io/

SUMMARY

Results-oriented Data Science and AI professional with over 3+ years of active experience in designing scalable data solutions, optimizing machine learning models, and building AI-driven applications. Proven skills in prompt engineering, ETL pipelines, and deep learning frameworks; strong track record of improving operational efficiency and providing actionable insights. Equally proficient in leveraging cloud platforms such as AWS and Google Cloud, big data technologies, and advanced visualization tools to ensure technical solutions are aligned with business objectives. Proactive leader with experience in working collaboratively within cross-functional teams to drive innovation and deliver measurable results. AWS and IBM Python for Data Science certified, with a Master's in Data Science from the University of Michigan.

WORK EXPERIENCE

Outlier Ai, Remote, USA **May 2024 to Present**
Ai Engineer / Prompt Engineer
Roles & Responsibilities:

- Designed high precision prompts to increase Ai model accuracy by 20% further enhancing relevance in key business applications.
- Partnered with data science teams to align prompt strategies with business objectives, driving impactful Ai output.
- Minimized biases and broadened adaptability through iteration of prompts to enable diverse use cases.
- Leveraged OpenAI and Hugging Face platforms to create solutions for focused prompts that place AI as a key driver of business value.
- Conducted prompt performance analysis through the analysis of key metrics, iterating on underperforming prompts, thus providing quantitative insights to drive refinement of prompts and model enhancements.
- Tested and optimized prompt variations to improve model interpretability and response accuracy, achieving a streamlined workflow and reducing response time by 15%.

Tata Consultancy Services, Bengaluru, India **October 2020 to July 2022**
Data Engineer – ETL Specialist
Roles & Responsibilities:

- Engineered advanced data mappings and transformations, optimized data processing from Teradata into PostgreSQL, and integrated data into Hadoop. This reduced the processing time by 30% and further increase the efficiency in the flow of data.
- Buit robust data validation protocols that increased the accuracy of analytical data by 25%, thus providing reliable insights.
- Scheduled and monitored jobs on mainframes. Efficiently identified and resolved errors that reduced operational downtime by 20% to secure daily operations continuously.
- Extracted data from different warehouses like PostgreSQL, Hive, and Teradata by writing optimized SQL queries.
- Wrote automation scripts and workflows in Shell and Python, applying them to the cleansing and transformation of data to reduce ETL processing time by 25%, hence improving the overall performance and reliability of the system.
- Spearheaded knowledge transfer sessions to onboard team members effectively about project tools and workflows.
- Closely collaborated with the business analysts, data scientists, and product managers to keep the ETL processes aligned with business requirements. Ensured data output was accurate and on time, corresponded with the project goals of.
- Conducted root cause analysis of recurring issues and did preventive measures, bringing down error rates by 30%.
- Automated regular data extraction which resulted in a reduction of manual effort by 40% hence improving data retrieval efficiency.
- Integrated advanced data validation techniques within the ETL process flows to much improve data quality and integrity for correct business insight.

Indian Servers, Vijayawada, India **May 2019 to September 2020**
Data Science Analyst
Roles & Responsibilities

- Supported and validated and impactful healthcare and agriculture data science projects using key **deep learning and computer vision techniques.**
- Contributed to the design of data pipelines in Cloud SQL and implemented data governance standards that improve the reliability of the data.
- **Played a key role in maintaining project quality** by participating in code reviews, sharing technical insights, and fostering collaboration, which contributed to smooth project execution and reliable outcomes.
- Translated business objectives into clear, actionable tasks: **created clear specifications, data mappings, and test cases**. Utilized **SQL in data analysis**, checking integrity, and aligning data outputs to project objectives both in healthcare and agriculture.
- Leveraged Python libraries like **NumPy, Pandas, and Matplotlib** to clean, map, and analyze data, supporting valuable insights through data preparation and feature engineering. Conducted data exploration in Python and R to enhance project accuracy.
- Collaborated effectively on data modelling and dashboarding using **Tableau** and **Power BI** by converting information into visual stories that best suit the interest of the clients in deriving meaningful insights, hence making optimized data presentation.
- Collaborated with data scientists in integrating **machine learning models** into the data workflows to drive project outcomes, enabling decision-informing predictive analytics.

TECHNICAL SKILLS

- **Programming:** Python, R, SQL, Java, Shell Scripting, HTML, JavaScript.
- **AI & ML**: Deep learning, NLP, Generative AI, LLMs, computer vision, predictive modelling (TensorFlow, PyTorch, OpenAI, Hugging Face).
- **Data Engineering**: ETL pipelines, Informatica, PostgreSQL, Teradata, Hadoop, Hive, Spark, Big Query.
- **Cloud:** AWS, GCP, Azure, Google Cloud SQL.
- **Visualization**: Tableau, Power BI, Seaborn, Matplotlib, Plotly.
- **Tools & Collaboration**: Git, Jira, CI/CD, Agile methodologies.
- **Operating Systems**: Windows, Linux, macOS.

EDUCATION

Master in Data Science | University of Michigan. Dearborn, Michigan, USA **August 2022 to April 2024**
Coursework Details: Big Data Analytics & Visualization, Deep Learning, Machine Learning, Neural Networks, Large Language Models, Database Management Systems, Artificial Intelligence, Applied Regression Analysis, Advanced Data Mining, Applied Data Analytics & Modeling for Enterprise Systems, Security and Privacy in Cloud Computing, Management Science.

B. Tech in Computer Science and Engineering | V R Siddhartha Engineering College, Vijayawada **June 2016 to September 2020**
Coursework Details: Data structures, Programming in C, Programming in Python, Object Oriented Programming Using Java, Computer Organization, Computer Networks, Database Systems, Operating Systems, Internet of Things, Design and Analysis of Algorithms, Probability and Statistics, Calculus, Complex Analysis and Numerical Methods, Image Processing, Cloud Computing, Cyber Security.

PROJECTS

Aqua Robot for Dead Shrimp Detection | *Python, TensorFlow, OpenCV*
Engineered an intelligent application using TensorFlow's Object Detection API to identify and monitor dead shrimp in aquatic environments with preciseness. This achieved a 98% improvement in data accuracy and empowered healthier aquatic health monitoring processes.

AI Resume Reviewer | *Python, Natural Language Processing, Stramlit, OpenAi API*
Developed an interactive application using Python, Streamlit, and OpenAI API to provide real-time feedback on resumes, optimizing their alignment with job descriptions. The tool incorporates Natural Language Processing (NLP) to analyze readability, formatting, and keyword relevance, offering actionable suggestions to enhance resumes for industry-specific roles. Designed a user-friendly interface and maintained the project on GitHub, implementing iterative improvements based on user feedback.

Chatbot Debate | *Python, NLP, Machine Learning, gradio*
Developed an AI-powered debate system featuring simulated discussions between personas like Steve Jobs and Elon Musk, utilizing GPT-4o and LLaMA-3.3 models via the Groq API. Implemented a referee bot to analyze transcripts and determine winners based on logic, innovation, and impact. Designed a dynamic chat interface for real-time debate visualization.

Genre Prediction App | *Python, NLP, Deep Learning, Streamlit*
Created a movie genre prediction application that classifies user-input descriptions as either Horror or Romance. Built using Streamlit, the app employs Word2Vec embeddings and a trained TensorFlow model for text representation and classification. Integrated SpaCy for text preprocessing to enhance prediction accuracy.

Diabetic Retinopathy Detection Tool | *Python, Deep Learning, Keras*
Led the development of an advanced deep learning application for detecting diabetic retinopathy from images of the retina, managing to achieve 90% diagnostic accuracy that will enable timely and correct medical interference in patients.

Advanced RAG Pipeline | *Python, Crawl4Ai, RAG, Chroma DB, Streamlit, Neural Networks, LLM's*
Developed a Retrieval-Augmented Generation (RAG) pipeline utilizing Chroma DB for efficient information retrieval. Implemented web crawling with Playwright and Crawl4AI to extract structured text and metadata from dynamic web pages. Performed semantic chunking and stored embeddings to facilitate rapid, contextually relevant responses. Deployed an interactive Streamlit web application to enable user-friendly search and retrieval functionalities.

AI Judge | *Python, Streamlit, Neural Networks, LLM's*
Designed a Gradio-based application to evaluate the performance of multiple AI models, including GPT-3.5 Turbo, GPT-4o, GPT-4, and GPT-4o Mini. The system simulates a judging process where each model alternates as a judge to assess responses from other models across predefined questions, aiming to identify the most accurate and relevant answers. Features include iterative judging, vote tracking, and automatic winner declaration.

CERTIFICATIONS

- **AWS**: Certified Cloud Practitioner
- **IBM**: Python for Data Science, Programming in Python
- **NPTEL**: | Programming, Data Structures, and Algorithms in Python; Database Management Systems