Saimanoj Bera
Saimanoj2509@gmail.com |+1 3134782240 | https://www.linkedin.com/in/saimanoj-bera-044831150/

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

- **Artificial Intelligence & Machine Learning:** Proficient in deep learning, computer vision, NLP, prompt engineering, and predictive modeling; experienced with model evaluation and tuning across algorithms (Logistic Regression, Random Forest, KNN, XGBoost, Gradient Boosting, etc.), neural networks, and large language models (LLMs) on platforms like OpenAI and Hugging Face; skilled with frameworks and libraries like TensorFlow, Keras, PyTorch, and NLTK.
- **Programming & Data Manipulation:** Strong proficiency in Python (NumPy, Pandas, Matplotlib, SciPy, Scikit-Learn), R, SQL, PL/SQL, Java, C, and Shell Scripting; familiar with NoSQL, HTML, JavaScript, XML, and tools like Jupyter Notebook, Anaconda, R Studio, and Google Colab.
- **Data Engineering & ETL**: Expertise in designing scalable ETL pipelines and automating data workflows, including data mapping, transformation, validation, and integration using tools like Google Cloud SQL, PostgreSQL, Teradata, Hadoop, and Hive; experienced in mainframe job scheduling and implementing data governance protocols.
- **Big Data & Cloud Technologies:** Hands-on experience with Google Cloud Platform (GCP), AWS, Azure, Storage Transfer Service, Cloud Scheduler, Hadoop, Apache Hive, Apache Spark, and Big Query, focusing on scalable data solutions and secure data storage.
- **Data Analysis & Visualization**: Advanced SQL skills for statistical analysis, time series analysis, customer segmentation, and market basket analysis (FP-growth, Apriori); adept at creating data-driven dashboards in Tableau and Power BI, with Python libraries like Seaborn, Plotly, and Matplotlib to visualize insights effectively.
- **Cloud Computing & Database Management**: Experience with Google Cloud SQL, AWS, and database management systems; focused on secure and efficient data storage, retrieval, and management within cloud-based infrastructures.
- **Project Collaboration & Process Improvement**: Strong background in Agile methodologies, cross-functional teamwork, knowledge sharing, and technical documentation; skilled at translating business objectives into actionable tasks and conducting code reviews to maintain high project standards.
- **Development Tools & Version Control:** Proficient with Git, GitHub, MS Excel, MS Word, MS PowerPoint, Adobe, and Putty, alongside collaboration and project management tools like Jira and CI/CD tools, ensuring efficient and organized workflows.
- **Operating Systems:** Comfortable working across Windows, Linux, and macOS environments.

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.

Predictive model Optimization for Hepatitis C | *Machine Learning, python*
For the best diagnosis of Hepatitis C, the optimized predictive accuracy is achieved through the evaluation and selection of the best-performing models: Logistic Regression, Random Forest, KNN, Decision Tree, Gaussian NB, XG Boost, and Gradient Boosting.

E-Commerce Customer Insights & Behavior Analysis | *Machine Learning*
Performed extensive customer behavior analysis with Time Series and KNN segmentation to uncover actionable insights. Performed Market Basket Analysis using FP-growth and Apriori algorithms to detect product associations driving targeted sales strategies and increased customer engagement.

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.

Data Driven E-commerce System | *SQL & Python, ETL*
Designed the e-commerce system, scalable in nature, using SQL and Python for efficient management of the ETL process on Google Cloud SQL. This enhanced storage and retrieval efficiency to a higher level, hence driving decision-making based on data.

Dashboard for AI- Generated Text Detection | *Python, Neural Networks, LLM's*
Designed an interactive dashboard that could identify the difference between human and AI-generated text, thereby paving ways for authenticity in the content across reliable differentiations between the text written by humans and that written by AI.

CERTIFICATIONS

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