SARIKA RAI

Pune, Maharashtra

| +91-9325463711 | sarikarai2904@gmail.com | linkedin.com/in/sarika-rai-93720a251/

SUMMARY

Results-driven Data Scientist with a strong foundation in statistics, machine learning, and real-time analytics. Proficient in Python, SQL, and PySpark with hands-on experience in time-series modeling, classification, and deep learning. Passionate about delivering data-driven insights and predictive solutions to solve complex business problems. Proficient in modern data workflows including ETL/ELT, real-time & batch processing.

Skilled in developing traditional ML models (SVM, Logistic Regression, GMM) and deep learning models (LSTM, CNNs, MLP). Hands-on experience with NLP applications and Generative AI frameworks. Well-versed with LangChain and OpenAI APIs for building Retrieval-Augmented Generation (RAG) pipelines and intelligent QnA systems. Proficient in prompt engineering techniques such as few-shot, zero-shot, ReAct, and Chain-of-Thought (CoT) prompting for enhancing model interpretability, response accuracy, and contextual understanding. Comfortable integrating structured and unstructured data sources.

EXPERIENCE

Data Engineer / Data Scientist
MaxGen Technology | Dec 2023 – Oct 2024

- Designed and implemented real-time and batch data pipelines using Python and PySpark.

- Built preprocessing workflows to support analytics and ML model development.

- Automated feature engineering and training pipelines using Apache NiFi.

- Developed dashboards and KPIs in collaboration with business teams to translate analytics into actionable insights.

- Integrated AWS and Snowflake for scalable storage and compute workflows.

Data Science Projects under MaxGen

RAG-Based Document Intelligence Assistant

- Built a Retrieval-Augmented Generation (RAG) pipeline using LangChain and OpenAI's GPT-3.5 Turbo to answer user queries from internal documents.

- Implemented document loaders, chunking, embedding (using OpenAIEmbeddings), vector storage in FAISS.

- Created a Streamlit UI allowing users to upload documents and ask context-aware questions.

- Applied advanced prompt engineering strategies such as few-shot, zero-shot, and ReAct prompting to improve relevance, coherence, and factual accuracy of model outputs.

Understanding Purchasing Patterns of an E-Commerce Giant

- Analyzed 100K+ e-commerce orders to uncover purchasing trends and customer segments.

- Wrote SQL queries using joins, CTEs, and window functions for product and demographic analysis.

- Suggested promotional strategies based on regional and seasonal buying behaviors.

Time-Series Based Terrain Identification Model

- Developed a terrain detection model using LSTM applied to accelerometer and gyroscope data.

- Achieved 91% accuracy through data normalization, sequence padding, and hyperparameter tuning.

Native and Non-Native English Speech Classification

- Built a classification model using MFCC features and models like GMM, kNN, SVM, and Neural Networks.

- Achieved 98% accuracy with NN and enhanced generalization using regularization.

Pyramid Image Blending (Computer Vision Project)

- Built a GUI-based image blending tool using Gaussian and Laplacian pyramids.

- Implemented from-scratch functions for convolution, downsampling, upsampling, and mask creation.

-

EDUCATION

Bachelor of Engineering in Artificial Intelligence
GH Raisoni College of Engineering | Aug 2021 – Apr 2024

CERTIFICATIONS

- Generative AI and LLM – Cisco (Expected Nov 2024)

- Certified Python for Data Science – Cisco (Expected Dec 2024)

- Machine Learning with R – Udemy (Expected Sep 2024)

SKILLS

Programming & Tools: Python, SQL, PySpark, JavaScript

Data Science & ML: Pandas, NumPy, Scikit-learn, TensorFlow, Keras, LSTM, SVM, Logistic Regression, GMM, Hyperparameter Tuning

Visualization & Reporting: Tableau, Power BI, Matplotlib, Seaborn

Big Data & Engineering: Apache Spark, Apache Kafka, Hadoop, AWS, Snowflake, Apache NiFi

Generative AI & NLP: LangChain, GPT-3.5 Turbo, OpenAI APIs, FAISS, Streamlit, Prompt Engineering (few-shot, zero-shot, ReAct, CoT)

Others: Git, GitHub, Microservices