Siddharth Dhaigude
siddharthdhaigude.me

Email : siddharthdhaigude@gmail.com
Mobile : +91-7506575930
Github: sd8capricon

About Me

I'm a passionate individual with a strong interest in Machine Learning and AI. Skilled in Python, with experience in Agentic AI, data processing, model development, and problem-solving. Eager to expand my expertise, I actively seek opportunities where I can leverage my skills to create impactful solutions.

Education

•Bachelor of Engineering in Computer Engineering (Hons. Data Science) *2020-24*
Don Bosco Institute of Technology, Mumbai - University of Mumbai CGPA: 8.83/10
Coursework: Data Structures & Algorithms, Relational Database, Machine Learning.

Experience

• Software Engineer II - Prorigo Software *Nov 2024 - Present*
- Led the design and development of a Graph-based Retrieval-Augmented Generation (RAG) system using Langchain & LangGraph, implementing modular adapters for Azure AI Foundry, and Ollama to integrate with ARAS and Protrak PLM platforms.
- Built an AI agent to automate BOM structure imports into ARAS, leveraging document parsing and entity extraction for accurate data mapping.
- Built a multimodal data ingestion pipeline using LlamaIndex, for extracting, processing, and transforming structured and unstructured content (text, images, tables) into a knowledge graphs for downstream semantic querying.
- Developed and deployed a FastAPI-based RESTful API exposing RAG and agent capabilities with robust authentication, scalability, and integration support.
- Owned end-to-end deployment, containerizing the solution and automating builds and deployments via CI/CD pipelines to Azure Container Registry and Azure VMs.

• Data Science Intern - Watershed Organisation Trust (WOTR) *Oct 2023 - Apr 2024*
- Responsible for processing and analyzing geospatial data, utilizing GIS systems like Google Earth Engine with the JavaScript API, to provide insights for informed decision-making.
- Developed and implemented a deep learning model in TensorFlow to predict wheat yield in the Gangapur region of Maharashtra, India, collaborating with local farmers to collect real-world data through surveys.
- Performed land use land cover (LULC) classification utilizing Random Forest and SVM algorithms to analyze and interpret land usage in the region.

Technical Skills

Languages: Python, Typescript, Javascript, HTML, CSS
Databases: Relational Database (SQL), Neo4j Graph Database
Libraries & Frameworks : Tensorflow, Numpy, Pandas, LLamaIndex, Langchain, LangGraph, FastAPI, Streamlit
Tools & Platforms: Git, Github, Docker, CI/CD Pipelines, Azure

Publications

S. Dhaigude, H. Panigrahi, S. Godse and P. Shaikh, "Smart Posture Analyzer For Exercise," 2023 6th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2023, pp. 389-394, DOI: 10.1109/ICAST59062.2023.10455069 - Link

Projects

• Smart Posture Analyzer for Exercise - GitHub
Android App to monitor workout accuracy
- Developed an android app to analyze the body posture in real-time
- Used OpenCV and Google's Mediapipe library to extract key body landmarks
- Implemented an on device LSTM based Recurrent Neural Network to process the body posture using Tensorflow
- Technology Used: Python, OpenCV, Numpy, Pandas, Tensorflow, Android SDK.

• Fine-Tuning Gemma 3 for Neo4j Cypher Query Generation - GitHub
Parameter-Efficient Fine-Tuning with QLoRA
- Fine-tuned the Gemma 3 model using QLoRA to generate accurate Neo4j Cypher queries from natural language prompts.
- Prepared and curated a custom dataset mapping natural language intents to Cypher query structures aligned with graph schema.
- Optimized training using parameter-efficient fine-tuning and **quantization** for resource-constrained environments.
- Technology Used: Python, QLoRA, Hugging Face Transformers, Neo4j.