DONDA SAI SIDDHU

+91 9959724308 ◇ Visakhapatnam, India
siddharthsiddhuw7@gmail.com ◇ linkedin.com/in/siddhu-donda ◇ github.com/siddhudonda

PROFILE

AI/ML Engineer with 1+ years of real-time experience in fine-tuning Large Language Models (LLMs) and building Retrieval-Augmented Generation (RAG) agents. Passionate about solving real-world problems using machine learning and deep learning. Proven expertise in delivering scalable AI solutions in agriculture, fraud detection, natural language understanding, and LLM applications.

EDUCATION

GITAM University, Visakhapatnam, India *2020 – 2024*
B.Tech in Computer Science and Engineering
CGPA: 7.75 / 10

FIITJEE College, Visakhapatnam, India *2018 – 2020*
Intermediate Education

TECHNICAL SKILLS

Languages	Python, Java
Databases	SQL, PL/SQL
Libraries	NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch, JAX
Frameworks	Hugging Face, LangChain, LangGraph, DSPy, LlamaIndex, Haystack
Machine Learning	Regression Models, XGBoost, Reranking, RL, LoRA, GRPO
Deep Learning	Vision Transformers, LLMs, Prompt Engineering
Vector DBs	FAISS, Pinecone, ChromaDB, Weaviate, Vertex AI Vector Search
Tools	Git, Docker, Flask, FastAPI, Gradio, MCP, n8n
Cloud	GCP (BigQuery, Vertex AI), AzureML, Gemini API

EXPERIENCE

Machine Learning Engineer *Feb 2023 – Dec 2024*
Farminno-Tech Pvt Ltd

- Developed and deployed real-time wildlife and insect detection models on Raspberry Pi 4 devices to enhance agricultural protection systems.
- Created an AI-based deterrent mechanism that triggers alarms and notifications upon threat detection; contributed to patent applications and DST funding proposals.
- Collaborated with hardware and firmware teams to optimize detection latency and reduce false positives in field conditions.

Data Science Intern *Sept 2023 – Dec 2023*
Oolook

- Built scalable scraping pipelines for Instagram, Facebook, and LinkedIn using Python and Selenium to gather gym and fitness influencer data.
- Performed advanced analytics on engagement metrics, follower-to-post ratios, and growth trends to identify high-performing profiles.
- Created visual dashboards and reports to help stakeholders make data-driven marketing decisions.

PROJECTS

- **Research QA System using LLM + Vector Embeddings:**
  - Built a scalable QA system for scientific research by embedding 250+ academic PDFs using the BAAI embedding model, enabling semantic search across dense research content.
  - Designed a hybrid retrieval architecture using Pinecone vector DB with GCP BigQuery (for metadata) and Cloud Storage (for PDFs).
  - Initially tested using Mistral v0.3 Instruct (7B parameters) before transitioning to Qwen-3 (4B) for improved accuracy and reduced latency.
  - Created a custom benchmark dataset of 6000+ domain-specific questions from PDFs and evaluated model performance using BERTScore.
  - Achieved an evaluation BERTScore of 85%, significantly improving domain-specific QA performance for researchers.
- **Fine-tuning DeepSeek-R1-Distill-LLaMA-8B (Medical LLM):**
  - Used HuggingFace and Unsloth with SFTTrainer to fine-tune the LLaMA-8B model for medical chain-of-thought reasoning.
  - Improved reasoning accuracy over baseline models using a custom supervised dataset.
- **Chat With Any PDF using RAG:**
  - Built a RAG-based PDF QA app with Gradio UI using FAISS for vector indexing and Google Gen AI for response generation.
  - Enabled semantic search and multi-document PDF analysis in a lightweight deployable interface.
- **LLaMA 3.1 8B Fine-Tuning Project:**
  - Fine-tuned Meta's LLaMA 3.1 8B model using Unsloth, HuggingFace, and LoRA for custom dataset adaptation.
  - Focused on improving instruction-following and factual reasoning using domain-specific medical and QA data.
  - Achieved notable performance gains in multilingual reasoning benchmarks.

PUBLICATIONS AND PATENTS

Publication: Click Fraud Detection in Online Ads using Ensemble Models — Published in AI-PTIS 2024.

Patent: AI-powered wildlife deterrent system using real-time threat detection and automated alerting — filed under Farminno-Tech Pvt Ltd with DST funding support.

ACHIEVEMENTS

1st Prize - Machine Learning Hackathon: Predicting Telecom Customer Behavior
Awarded 1st place among 150+ participants for designing a highly accurate ML model to predict customer churn using classification algorithms and ensemble methods. Presented model insights to a jury of industry experts, highlighting data-driven retention strategies.