

Vadlamudi Saketh Sri Harsha

Hyderabad | 9704808576 | saketh22off@gmail.com | [LinkedIn](#) | [GitHub](#) | [LeetCode](#) | [Website](#)

PROFESSIONAL WORK EXPERIENCE (3.7 Years)

TIGER ANALYTICS- AIML Senior Associate

May /2025 - Present (Hyderabad)

- Architected an autonomous, **multi-agent** reasoning engine using LangGraph to decompose ambiguous user requests into parallel, dependency-aware execution graphs, **reducing query latency by 60%**.
- Engineered the engine as a **self-correcting, schema-aware RAG pipeline** where orchestrated agents autonomously produce **natural language insights, structured tabular data, and interactive Chart.js configurations** from a single user query.
- Tech Stack : Python, LangGraph, Pydantic, AWS (Lambda, DynamoDB, RedShift, Bedrock), Fast API

VALUELABS - Data Scientist || (Senior Software Engineer)

Mar/2022 – May/2025 (Hyderabad)

- Developed a Machine Learning (ML) model for **Sentiment Analysis**, which accurately classifies customer reviews on products. Achieved 93% accuracy; insights were adopted by product stakeholders to refine customer messaging.
- Architected an **Advertisement Recommendation Platform** using collaborative filtering, achieving MAE of 0.79 and improving ad relevance, leading to a 20% increase in click-through rates.
- Deployed ML models using AWS services such as **EC2, ECR and SageMaker**, leveraging **Docker** for seamless containerization and scalability, while using reliability through robust cloud Infrastructure.
- Spearheaded **Python processes on Factory Design Mechanism**, reduced run time performance by 2x and 12% decrease in Campaign drop off rate.
- Implemented several **Python-based automations** for handling manual/regular tasks resulted in significant reductions of over 65 working hours monthly with notable recognition from stakeholders.
- Delivered 4 high-quality, scalable, and secure **Web applications using Django** framework by creating innovative, user-centric solutions that have consistently exceeded client expectations.

ADDITIONAL WORK

GEN AI : - HybridSeek ([Link](#)) – Semantic RAG Search Engine

Tech Stack: Python, FastAPI, OpenSearch, Redis, Langfuse, Ollama, Jina AI, Airflow, Docker

- Built a hybrid RAG pipeline (**BM25 + dense vector + RRF**) for searching over **10k+ document chunks**.
- Designed a **section-aware chunking + 1024-dim embedding pipeline**, achieving **>90% semantic recall@10** during qualitative evaluation.
- Reduced repeated query latency by **~8x using Redis caching** with exact-match hashing.
- Automated ingestion and re-indexing with **Airflow DAGs**, processing **1k+ new papers/day**.
- Integrated **streaming LLM inference with Ollama**, delivering **sub-4s first-token latency**.

NATURAL LANGUAGE PROCESSING:

- Language Modelling ([Link](#)): Engineered a state-of-the-art language model using pre-trained advanced transformers like DistilBert, Bert2, fine-tuned on custom dataset for accurate prediction of top 10 masked word and seamless contextually rich text generation.

EDUCATION

Executive **Post Graduate** Certification in Data Science and AI from **IIT Roorkee**

2024 - 2025

B.Tech in CSE from Narasaraopeta Engineering College (8.94 CGPA)

2018 - 2022

AWARDS & CERTIFICATIONS

- ValueLabs CEO Choice Award Winner 2024** ([Link](#))
- GCP Professional ML Engineer, GCP AI Engineer For Leaders ([Link](#)), AWS Cloud Practitioner ([Link](#))
- AI for Everyone ([Link](#)), Python Data Structures ([Link](#)), Programming for Everybody ([Link](#)) (From Coursera)

SKILLS

- Languages : Python, SQL
- AI/ML: Machine Learning, Natural Language Processing (NLP), Transformers , Generative AI, Retrieval-Augmented Generation (RAG) , LangChain, LangGraph
- Cloud Platforms: AWS (Lambda, DynamoDB, Redshift, EC2, ECR, SageMaker, S3)
- Tools : Git, Unix, Docker, Jira, FastAPI