Surya Teja Menta

Data Scientist / ML Engineer / AI Engineer

☎ +91 8309584461   @ mentasuryateja   ⚬ Surya-Teja-Menta   in surya-teja-menta   ⊕ suryatejamenta.co.in

Address: 3-12-16, Kotha Bazar, Kavali (524201), Andhra Pradesh, India

Summary

I'm <u>Surya Teja Menta</u>, Results-driven Senior Data Scientist with 4+ years of experience in **Data Science, Machine Learning (ML), and Generative AI (GenAI)**. Proven expertise in **RAG (Retrieval-Augmented Generation), LLM fine-tuning,** MLOps, and end-to-end AI solutions. Strong background in data analytics, statistical modeling, AI-powered automation, and scalable AI architectures. **IBM Certified Professional Data Scientist** with hands-on experience in **LangChain, Hugging Face, OpenAI APIs**, **Vector Databases** (ChromaDB, Pinecone), and cloud deployments (AWS, GCP). Passionate about AI research, model optimization, and developing cutting-edge AI solutions.

Work Experience

- Senior Data Scientist - Robert Bosch (Jan 2024 - preset)

Designed and deployed **custom YOLOv8 models** for computer vision applications. Built **multi-modal RAG pipelines** for intelligent **document processing, contextual retrieval, and GenAI-powered insights.** Worked on **Time series data** for **Adnoc** Project. Developed ML algorithms for signal processing and sensor vibration data analytics using scikit-learn, PyTorch, and TensorFlow. Worked on **LLM fine-tuning, prompt engineering**, and efficient model inference for enterprise AI applications. Integrated MLOps best practices for scalable AI pipelines, model versioning, and automated deployment.

- Subject Matter Expert (Data Analyst) - <u>Tudip Technologies</u> (Oct 2020 - Oct 2023)

Developed **interactive analytics dashboards** and data reports in Google Data Studio for content management analytics. **Automated text summarization & paraphrasing using NLP Transformers (BART, T5)** and deployed on **Google Cloud** Run. Built AI-powered data visualization tools to enhance real-time business intelligence and trend analysis. Integrated **LLM-powered chatbots** with Google Sheets API to improve data accessibility across teams. Recognized with the **Best Team Award** for highest client appreciation and impactful AI-driven automation.

Projects

- **Advanced RAG Research Assistant (LangChain, LLM, Hugging Face, Vector DB, AI, NLP)**
  Built an AI-powered **Retrieval-Augmented Generation (RAG) system** for intelligent document analysis
  with multi-modal PDF processing, semantic chunking, and **adaptive query handling**. Integrated
  vector-based **semantic search (ChromaDB), BM25 retrieval**, and contextual response generation using
  Llama3.2 and Hugging Face. Developed context-tracking algorithms for personalized, stateful interactions.
  Tech stack**: Python, LangChain, Chainlit, OpenAI API**.

- **Github Automated Analysis (OpenAI, Langchain, LLM)**
  Automated Complexity Analysis: Engineered a Python tool using GPT and LangChain to pinpoint intricate GitHub
  repositories. **Memory Optimization:** Implemented efficient memory management for processing extensive files while
  adhering to token limits. **Enhanced Complexity Scoring:** Utilized prompt engineering to elevate accuracy in
  evaluating repository complexity, refining the precision of the tool.

- **Re-Enhance.AI (Opencv, ESRGAN, Pytorch)**
  The Re-Enhance.AI project is a set of tools and algorithms that can be used to improve the quality of your
  Image for Space & **Research Purposes**. Even though the model accepts only images with (256x256x3) dimensions,
  this framework manages the higher dimensions by **Split and Send Policy**.