PANCHADA VAMSI

+91-7997115484 ◇ India
vamsivs58@gmail.com ◇ linkedin.com/vamsi ◇ github.com/vamsi

OBJECTIVE

Software engineer with 1+ years of experience in a product-based company, skilled in designing LLM tools, software modules, packages, agents, data pipelines, and REST APIs.

EDUCATION

Bachelor of Information and Communication Technology, SASTRA University 2019 - 2023

TECHNICAL SKILLS

- **Programming Languages:** Python (Advanced), Scala, SQL, Java
- **AI/ML Frameworks:** TensorFlow, PyTorch, Hugging Face Transformers, Scikit-learn, XGBoost
- **Generative AI Tools:** OpenAI GPT, NVIDIA NIM, Llama3, Ollama, Sentence Transformers
- **Big Data Technologies:** Apache Spark, Airflow, Kafka, Hadoop, ETL Pipelines
- **Cloud Platforms:** AWS (S3, EMR, EC2), Azure (AKS, Synapse), Kubernetes
- **Backend Frameworks:** FastAPI, Flask, Django, Spring Boot
- **Databases:** PostgreSQL, MySQL, ChromaDB, FAISS, Elasticsearch
- **DevOps Tools:** Docker, Kubernetes, CI/CD Pipelines
- **Data Visualization:** Matplotlib, Seaborn, Plotly

EXPERIENCE

Associate Engineer Aug 2023 - Present
Crayon Data *Chennai, India*

- Optimized large-scale data processing systems using Apache Spark, Airflow, Scala, and SQL, improving performance for banking clients including HDFC and ADIB in the Offer Marketplace.
- Enhanced and maintained offer management platforms with advanced features such as pattern matching and cross-border data publishing to Elasticsearch.
- Developed data validation and mapping solutions utilizing NLP, vector databases (Chroma DB, FAISS), and LLMs, significantly improving data accuracy and processing efficiency.
- Designed and optimized ETL pipelines and data quality frameworks using Spark SQL, DAGs, and machine learning techniques for automated rule generation in a centralized data warehouse.
- Implemented secure, scalable microservices and APIs with FastAPI, Docker, and NVIDIA Inference Microservice (NIM) on Azure Kubernetes Service (AKS), automating model deployment via Apache Airflow DAGs.
- Created frameworks for LLM response parsing with regex, and developed DAGs for language model deployment; implemented DQL-only filtering to improve turnaround time by 30
- Built a business validation system powered by LLMs to ensure the integrity of business data and improve overall data reliability.
- Developed multiple proof-of-concept applications utilizing Generative AI, Agents, and Memory systems for personalized lifestyle and financial applications.
- Created and supported common modules for multiple LLM applications, improving reusability and efficiency across various projects.

- Developed an Agent Controller and Runner to manage and automate agent-based workflows, enhancing scalability and coordination of tasks across LLM-driven applications.

PROJECTS

Customer Call Analytics Platform. Developed a Streamlit application integrating OpenAI Whisper and GPT-4 for transcription, sentiment analysis, and actionable insights, reducing manual review effort.

Fish Weight Prediction System. Built an ML pipeline for predicting fish weight using XGBoost, deployed with Streamlit for real-time use.

FastAPI-Based Language Model API. Designed a FastAPI-based solution with multi-threading and Docker integration for high-performance NLP tasks.

Pneumonia Detection System. Created a deep learning model using CNNs for detecting pneumonia from X-ray images, achieving high accuracy.

ACHIEVEMENTS & CERTIFICATIONS

- Certified Oracle Cloud Infrastructure 2024 Generative AI Professional
- Published Research Paper: Grid Sampling Based Hypergraph Matching for Multiple Object Tracking
- Published Research Paper: Enhancing Pneumonia Detection with Masked Neural Networks
- Recipient of H1 Leap Award for Outstanding Generative AI Innovation
- Active contributor to technical blogs and community discussions on AI/ML career development