

Sasank Jampana

+91 9347004843 | sasankjampana8@gmail.com | linkedin.com/sasankjampana | github.com/sasankjampana8

Profile

- 3 years of experience delivering ML and LLM products from research to production. I focus on fast experimentation, strong evaluation, and dependable deployment.

Experience

Intellect Design Arena

June 2023 – Present

Hyderabad, Telangana

Data Scientist

- Building an Agent to create dynamic summaries for different customer dashboard sections. Utilizes web search to provide custom recommendations
- Built an in-app RAG assistant for the Wealth application's product support that provides step-by-step guidance, feature discovery, and annotated screenshots; indexed product docs and release notes; integrated feedback loops to reduce repetitive tickets and accelerate onboarding.
- Implemented a batch RFP analyzer using a RAG agent that ingests multiple proposals, evaluates availability vs. gaps, and returns per-proposal confidence scores; streamlined sales triage and shortened response cycles.
- Developing a multi-agent Wealth EKG that answers from documents and live web data, generates visual analytics, and automates financial report ingestion, metadata creation via LLMs, and indexing for instant retrieval.
- Built a ReAct-based multi-agent chatbot with Oracle Text-to-SQL; integrated FastAPI with PII masking and AWS S3-powered chart visualizations for secure financial analytics.
- Orchestrated a full-stack multi-agent LLM system combining RAG (Claude 3.5), GPT-4o-driven Oracle Text-to-SQL, secure FastAPI retrieval, AWS S3 renderings, and dynamic knowledge/dataset lifecycle management.
- Developing a lightweight on-prem ML framework enabling rapid training, deployment, and monitoring of models with 90% less manual effort, distributed as a plug-and-play Python package.
- Designed predictive models including a product propensity model (stacked ensemble, 96.2% macro F1) and a churn prediction model (99% weighted F1), integrating SHAP explainability, drift detection, and deployment across AWS SageMaker, Docker, and on-prem environments.
- Achieved 95%+ accurate targeted extraction of key financial terms from market outlook reports using PyMuPDF and Difflib.
- Delivered 99% accurate CAS parsing in under 3 seconds using a custom regex engine.
- Reduced manual processing by 80%+ through LayoutLM, BERT, and OCR-based document classification and clustering.

Forbes Advisor

Dec 2022 – May 2023

Chennai, Tamil Nadu

Data Engineer Intern

- Built a real-time data pipeline PoC on GCP, integrating and validating diverse sources before ingesting into BigQuery; improved processing speed and data quality for analytics.

Projects

Emotion Classifier — Python, PyTorch, Transformers, Streamlit — GitHub — Demo

- Pre-trained a deberta base model on emotion text dataset to classify five different emotions. Scored 88.9 macro f1 score

Technical Skills

Languages & Frameworks: Python, SQL, TensorFlow, PyTorch, Scikit-learn, Hugging Face Transformers

Generative AI & LLMs: AWS Bedrock, LangChain, LangGraph, LangSmith, Prompt Engineering, RAG, PEFT/LoRA, Instruction Tuning

Data & Databases: PostgreSQL, MySQL, MongoDB, PGVector, Pinecone, FAISS

MLOps & Tools: AWS (SageMaker, Lambda, S3, ECR, SNS), GCP (BigQuery), Docker, MLflow, FastAPI, Git, GitHub

Education

Vellore Institute of Technology, Andhra Pradesh

July 2019 - May 2023

Bachelor of Technology in Computer Science (CGPA: 8.93 / 10.00)

Amaravati, AP

Certifications & Achievements

Diploma in Data Science, IIT Madras

2x Spotlight Award: Recognized for exceptional contributions to the Customer Engagement Indicator project and PlanMyDay usecase, acknowledged as a valuable resource for driving the project's success and delivering impactful results.