

Panchada Vamsi | AI and Data Engineer

📍 Visakhapatnam ,Andhra Pradesh 📞 +91 7997115484 ✉ vamsivs58@gmail.com
in /panchada-vamsi 🌐 /vamsi-panchada </> leetcode.com/panchada-vamsi </> huggingface.co/vamsiPanchada

Certified Oracle Generative AI Professional with one plus years of experience in building and deploying large language models (LLM) and **spark** based data pipelines to automate processes in **airflow**. Experienced in data validation and schema validation to ensure data integrity and quality. Strong foundation in **Gen AI**, gained through a B.Tech. from SASTRA University, with a focus on applying generative AI solutions to improve efficiency in real-world business tasks.

Education

Program	Institution/Board	%/CGPA	Year
Bachelor's degree	SASTRA University, Thanjavur	8.01	2023
12th	Board of Intermediate Education, AP	96.7%	2019
10th	Secondary School Certificate, AP	9.7	2017

Work Experience

1. Associate Engineer Aug 2023 - Present
Crayon Data (Chennai)

- Optimized large-scale data processing systems using **Apache Spark, Airflow, Scala, and SQL** improving performance for banking sector clients including HDFC and ADIB for Offer marketplace.
link to demo
- Enhanced and maintained offer management platforms, optimising advanced features like **pattern matching** and **cross-border** data publishing to **Elasticsearch**.
- Developed innovative data validation and mapping solutions using **NLP, vector databases (Chroma DB, FAISS)**, and **LLMs**, significantly improving accuracy and efficiency.
- Created and optimized **ETL pipelines** and **data quality** frameworks, utilizing **Spark SQL, DAGs**, and machine learning techniques for an advanced automated rule generation for a centralised **data warehouse system**.
- Implemented secure, scalable microservices and APIs using **FastAPI, Docker**, and NVIDIA Inference Microservice (**NIM**) for on-premises model deployment using Azure Kubernetes Service (**AKS**) and automated the process using a dag.
- Created frameworks for LLM response parsing using **regex**, and for Language Model deployment via DAGs and implemented DQL-only filtering for database protection that brings down the turnaround time to 30% and increases the accuracy over 40% than **Ollama**.
- Technologies:** Apache Spark, Apache Airflow, Python, SQL, Elasticsearch, NLP, Machine Learning, FastAPI, Docker, ETL, Big Data, Sentence transformers, Ollama, meta/llama3

Key Projects

1. Fish Weight Prediction System

- Developed a machine learning pipeline for predicting fish weight using **XGBoost, Decision Tree**, and **Random Forest** models.
- Preprocessed the Fish Market Dataset by cleaning data and applying **Label Encoding** for categorical variables.
- Performed model evaluation using **MSE, MAE, and R² score**; optimized **XGBoost** with **HyperOpt** for better accuracy.
- Deployed an interactive web app using **Streamlit**, allowing users to input fish features and receive real-time weight predictions.
- Leveraged **Pandas, NumPy, Scikit-learn, and Matplotlib** for data handling, model building, and visualization.
link to demo
- Technologies:** Machine Learning, Hyperparameter Tuning, Data Preprocessing, Streamlit Deployment, XGBoost

2. Customer Call Analytics Platform

- Developed a **Streamlit** web application for analyzing customer call recordings and generating insights.
- Utilized **OpenAI Whisper** for automated transcription of audio files in formats like mp3, wav, and ogg.
- Integrated **GPT-4** to provide real-time call summaries, conversation analysis, and insights.
- Implemented a **Chroma vector database** for query-based transcription searches and retrieval.
- Processed and visualized customer call data using **Python Pandas** to generate daily, weekly, and monthly summaries.
- Extracted key metrics, including customer sentiment, feedback, and final decision status.
- Built a chat interface for interactive querying of call data and **AI-powered** insights.
- Reduced manual effort and improved accuracy in reviewing call transcripts and customer feedback.
- **Technologies:** Python, Streamlit, OpenAI GPT-4, Whisper, Chroma (Vector Database), Pandas

3. FastAPI-Based Language Model API with Docker Integration

- Developed a **FastAPI-based application** for serving a language model using the **ctransformers** library to handle natural language processing tasks.
- Implemented a **Pydantic** schema for validating API requests, ensuring that input data adheres to the expected format.
- Configured **Docker container** to encapsulate the **Python** environment and dependencies, enabling consistent deployment and scalability.
- Optimized model performance with **multi-threading** support and a token limit configuration to manage computational resources efficiently.
- Integrated **Uvicorn** as an **ASGI** server to serve the **FastAPI** application, providing asynchronous support and high performance.
link to demo
- **Technologies:** Python 3.9, FastAPI, Pydantic, Docker, Uvicorn, ctransformers

Course Projects

1. Enhancing pneumonia detection with masked neural networks

2023

Smt. L.Gowri, Team Size: 3

deep learning approach

- Built a Streamlit-based web application to detect pneumonia from chest X-ray images using **deep learning** models (**UNet** for image segmentation and custom **CNN** for classification).
- Integrated TensorFlow and **Keras** for model development, with **OpenCV** for image preprocessing and visualization.
- Optimized performance using **streamlit** cache resources and implemented K-Nearest Neighbors (**KNN**) for final classification.
- Enabled file upload, mask visualization, and segmentation display for enhanced user interaction.
link to demo
- **Technologies:** Python, TensorFlow, Keras, OpenCV, Streamlit, NumPy, Pickle.

Online Courses

- Become an OCI Generative AI Professional
- Generative AI Courses [Coursera]
 - Generative AI for Everyone (Andrew Ng)
 - Fundamentals of Generative AI for Beginners
 - Generative AI with Vertex AI: Build a customer chatbot
 - LangChain Chat with Your Data
 - Functions, Tools and Agents with LangChain
- Data Engineering Essentials using SQL, Python, and PySpark (Udemy)
- Building Real-Time REST APIs with Java Spring Boot (Udemy)

Technical Skills

- **Programming Languages:** Python, Scala, SQL, Java
- **Big Data Technologies:** Pyspark, Scala Spark, Spark sql, Kafka, Hadoop, Airflow, Zepline
- **Gen AI:** Ollama, NIM, Hugging Face, transformers, Vector DB (Chroma DB, FAISS), pytorch, S-BERT
- **Backend Technologies:** FastAPI, Django, Spring Boot, Flask, pydantic, uvicorn
- **DevOps:** Docker, Kubernetes
- **Big data Cloud:** AWS s3, AWS EMR, AWS EC2, Azure Virtual Machine, Azure Data Lake Service, Synapse Analytics
- **Machine Learning:** Pandas, Numpy, NetworkX, Spark ML, scikit learn, plotly, seaborn, Streamlit, keras, tensorflow, open-cv
- **API Development:** RESTful APIs
- **Database Technologies:** SQL, PostgreSQL, MySQL, Chroma, FAISS
- **Version Control:** Git (GitHub, Bitbucket)

Achievements

- Published Research Paper on Grid sampling based hypergraph matching technique for multiple objects tracking in video frames
[link to paper](#)
- Published Research Paper on Enhancing pneumonia detection with masked neural networks: a deep learning approach
[link to paper](#)
- Recognized as a quick starter, awarded the H1 Leap Award within the first year for outstanding contributions to projects and innovation in generative AI.
[link to post](#)
- Certified with Oracle Cloud Infrastructure 2024 Generative **AI Certified Professional** Exam Number: 1Z0-1127-24 312936027OCI2024GAIOCP