Pavan Kumar Balijepalli

+91 7673977636
2021sc04115@alumni.bits-pilani.ac.in

EXPERIENCE

Senior Data Scientist — *Amnet Digital,* *Hyderabad*

SEPT 2023 - PRESENT

Worked for an e-commerce client and developed multiple generative AI systems using Prompt Engineering, LLM Output Structuring, and Agentic Orchestration that leverage the power of Large Language Models.

Responsibilities:

- Developed a Multi-Agent Orchestrated workflow using OpenAI 4.1 mini that can close user requests by categorising it and providing necessary information. Built on LangGraph, Azure AI Foundry Frameworks.
- Developed a custom Framework on top of Foundry SDK for ease of usage and re-usability.
- Built an e-commerce chatbot (similar to Amazon Rufus) using LangChain and GPT-4o, which can perform sales Q&A on the e-commerce DB. Implemented AI-based automation for both context compression and product retrieval. Finalised everything with Streamlit Application.
- Build a framework to develop and compare the GPT3.5 NL2SQL results with the NL2SQL leader, Defog SQLCoder, specific for the e-commerce domain.
- Fine-tuned Microsoft's Phi-2, the small language model which outperformed Mistral-7,13B for the NL2SQL use case.

Senior Machine Learning Engineer — *Nagarro,* *Hyderabad*

NOV 2022 - SEPT 2023

Worked for a food and beverage client and developed an anomaly detection model to ensure the quality of products in the production chain.

Responsibilities:

- Collaborated in the design of machine learning systems.
- Built an OCR look-alike model, using TensorFlow, which found anomalies in the products and ensured high accuracy without compromising time.
- Understood business objectives and developed models that help to achieve them, along with metrics to track their progress.
- Managed resources such as hardware, data, and personnel to meet deadlines.
- Analysed the machine learning algorithms that could be used to solve a given problem and ranked them by their success probability.
- Explored and visualised data to gain an understanding of it.

KEY SKILLS

Languages and Tools: Python, SQL, R, Excel (Pivot Tables, VLOOKUP), JupyterLab, Git, Toad

Data Processing: NumPy, Pandas, Seaborn, OpenCV

Statistical Analysis: Hypothesis Testing, Confidence Intervals

Machine Learning Algorithms: Regression and Classification, Neural Networks, Convolutions, Sequence Models

Model Development: TensorFlow, Keras, Sci-Kit Learn

Gen-AI Libraries: Transformers, BitsAndBytes, SentencePiece, PEFT, Datasets, Accelerate, TRL

Closed Source LLMs: OpenAI, Anthropic, Gemini, Mistral Large

Open Source LLMs: Phi, Mistral, Gemma, Deepseek, Qwen, Mixed-Bread, Llama

Agentic Libraries: CrewAI, SmolAgents, Ollama, LangChain, LlamaIndex, UnslothAI

App Development: Gradio, StreamLit

Cloud: Azure

Backend: Node.js, Flask API, Postman

Data Science Analyst — *Tiger Analytics*, *Hyderabad*

NOV 2021 - NOV 2022

Worked for an Insurance Client and developed a classification model to group customer data to process their profiles through the business rules of the insurance application.

Responsibilities:

- Built a classification model for customer data based on the International Classification of Diseases to predict if the customer's application for insurance should be approved.
- Automated a Model Health Monitoring Framework which monitored over 160 models and reported their performance, including the data drift, model run, and execution.
- Performed Mann-Kendall test on a Real-Time model for the International Classification of Diseases (ICD) on the customer data, which helped understand the trend between two time periods.
- Built Scripts to automate the onboarding of Models into the framework, modifying the change of the Benchmark period of models and raising incidents during a breach of performance thresholds.

Project Engineer — *Wipro Ltd.*, *Bengaluru*

JUN 2019 - NOV 2021

Worked for an Automotive Client and developed a Car Spare-part Availability Tracking System for service centres. Improved overall working accuracy and efficiency of the project using Machine Learning and Deep Learning techniques.

Responsibilities:

- Used various statistical and analytical methods to build a predictive Convolutional Siamese model which compares images from the service centre and from the automobile warehouse to recognise the availability of spare parts.
- Participate in the entire application lifecycle, focusing on coding and debugging.
- Integrated the model with the existing project pipeline and built a Flask API to perform requests from the backend.
- Containerised the Model with Docker, deployed it on the AWS server and collaborated with app developers to integrate user-facing elements with server-side logic.

EDUCATION

Master's of Technology (Data Science & Engineering)
from Birla Institute of Technological Sciences, Pilani
2024 - 7.9 CGPA

Bachelor's of Technology (Electrical and Electronics)
from Jawaharlal Nehru Technological University, Hyd
2019 - 8.21 CGPA

ACHIEVEMENTS

Fine-tuned 2 LLMs on my Local Machine over RTX 4070

Fine-tuned 7+ LLMs on Azure and Provider Instances

Dissertation Paper on Finetuned-Phi2 vs Defog Models

Personal Vibe Coding Agent on Qwen2.5-Coder for personal usage

LANGUAGES

English, Telugu, Hindi