

PAVAN KUMAR BALIJEPALLI

Hyderabad, India | pavanbalijepalli.bits@gmail.com | github.com/pavankumarbalijepalli

PROFILE SUMMARY

Data Scientist and AI Engineer with 6+ years of experience building and deploying ML and Generative AI solutions. Skilled in Python, LLMs, LangChain, and cloud deployment. Passionate about developing intelligent, scalable systems that merge data, automation, and user impact through applied machine learning and model optimization.

EDUCATION

Birla Institute of Technology and Sciences, Pilani

May, 2024

Masters in Data Science and Engineering

Relevant Coursework: 4 Semesters

CGPA: 7.94

Jawaharlal Nehru Technological University, Hyderabad

Apr, 2019

Bachelors in Electrical and Electronics Engineering

Relevant Coursework: 8 Semesters

CGPA: 8.21

WORK EXPERIENCE

Amnet Digital, Hyderabad, Telangana

Sept-2023 till Present

Senior Data Scientist

- Implemented Ticket Assistant as Friction Elimination which saved 8000+ human hours per month.
- Improved LLM Output Structuring which led to increase in Positive Agent Feedback by 250%.
- Developed in-house Q&A Agent which led to 70% fewer clicks by customer in dashboards.

Nagarro, Senior Machine Learning Engineer, Hyderabad, Telangana

Nov-2022 till Sept-2023

- Built an STT AI Kiosk which improved the night sales on Drive-thru by 30%
- Developed a Product Quality Classifier which led a successful Pilot.

Tiger Analytics, Data Analyst, Hyderabad, Telangana

Nov-2021 till Nov-2022

- Built a Model Monitoring Framework which provided KPIs for existing models to evaluate performance.
- Developed an Insurance application classifier which reduced high-risk insurance applications by 60% and saved multiple manual hours.

Wipro, Project Engineer, Bangalore, Karnataka

Jun-2019 till Nov-2021

- Worked on Quote Completion Classifier. Increased discounts for borderline customers to convert quotes to sales.
- Developed Sales Run Rate Forecasting Model which provided clear-sight to business on priorities over coming quarters.

PROJECTS

Service-Now Ticket Assistant for E-commerce Client

Amnet Digital

- Each Month 1200+ Look-up Quick cases are closed by support agents manually, which needs to be automated.
- Used a Multi-Agent Orchestration to achieve this, with each agent handling different kind of tickets.
- Pilot Phase itself saved 8000+ hours on a small category of tickets which led to further discovery and enhancement.

Semantic Product Search Agent for E-commerce Client

Amnet Digital

- Customers searching for long time to reach the right product when queried with filter based input.
- Used NLP + GenAI Agent to handle user input and query the DB. Implemented intent understanding on user question.
- Successful POC led to Standalone Feature. As per click-stream data, the AI reduced the amount of clicks by 40%.

Q&A Sales Agent for In-House Dashboard

Amnet Digital

- Customers exploring dashboard with multiple filters manually, and requested for chat-bot support for easy querying.
- Leveraged Open-AI Endpoints to prepare DB Agent which can query based on Natural Language Input.
- After making few modifications as per customer's requests, the release led to lower click-stream and higher thumbs-up rate on the chat-bot

Comparing Fine-tuned Phi-2 with Defog SQLCoder for Dissertation

BITS Pilani

- Fine-tuned a SLM locally on RTX 4070 which is further evaluated with Test Dataset for multiple runs.
- Evaluated and prepared a comparative analysis between Defog-7b Model vs Phi-2.
- Model is publicly available on HuggingFace and has 8,800+ downloads.

Speech2Text AI Kiosk for F&B Client

Nagarro

- Late night Stores with drive-thru are facing issue with Order mismatching due to human error.
- Using Open-AI Whisper along with RASA to build a complex hierarchical conversational chatbot.
- Increased sales by 30% due to chat bot suggesting add-ons to the customers in first 3M pilot phase.

Product Quality Classification for F&B Client

Nagarro

- Conveyor belt processing unclean products along with clean ones, needed automated surveillance and control.
- An Image Classifier on a Video Stream which classified the products into clean and unclean at 27 FPS.
- It led a successful Pilot with 97% accuracy on real time data.

Model Health Monitoring Framework for Insurance Client

Tiger Analytics

- Multiple models on the production deployment had a reduced performance over 8-9 Month window.
- A structured framework which is automated to fetch the output details for every run to find PSI and other KPIs.
- Understanding change in data distribution led to model re-training which led to consistent results over the year.

Insurance Risk Classification for Insurance Client

Tiger Analytics

- Applications are manually validated for high risk customers to reject the application to avoid insurance payments.
- Using I.C.D from W.H.O, customer data is prepared with mortality rate and a XGBoost Classifier is developed.
- Reduced manual validation by 80% by maintaining precision of 99% on real-time data.

Quote Conversion for Automotive Client

Wipro

- Predicting the chances of a quote to be converted into a sale by leveraging the customer data.
- Customer data is first binned using clustering for feature engineering, and then developed a linear classifier.
- The Pilot phase was successful with 94% production accuracy and reduced manual by 40%

Sales Run Rate Forecasting for Automotive Client

Wipro

- Forecasting the upcoming quarters by the amount of sales done by each seller for clear road-map into future
- Sales data is prepared for SARIMA Auto Regression model for each region and seller with 94% accuracy.

TECHNICAL SKILLS

- *Languages and Tools:*
 - Python, SQL, Excel, Git
- *Data Processing:*
 - NumPy, Pandas, Seaborn, OpenCV
- *Mathematical Foundations:*
 - Linear Algebra, Calculus, Probability, Statistics
- *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, Ollama, Unsloth
- *Agentic Frameworks:*
 - SmolAgents, LangChain, LangGraph, LlamaIndex
- *Cloud:*
 - Azure, AWS

PERSONAL PROJECTS

- *Finetuning Small Language Models:*
 - Fine-tuned 2 LLMs on my Local Machine over RTX 4070
 - Fine-tuned 7+ LLMs on Azure and Provider Instances
- *Research:*
 - Pre-trained a 20M SLM locally on HumanEval.
 - Fine-tuned 9B LLM for Telugu Language Understanding.
 - Reading HuggingFace Papers, and Arxiv.
 - Exploring cross industry breakthroughs.
- *Agentic AI Projects:*
 - K.U.W.T Agent: Agent which sends daily summarized updated on new technology in selected topic.
 - Review Retriever: Agent which connects to Google Search to find reviews on a product and summarizes it.
 - Second Brain: Agent upon feeding personal documents, stores it locally and answers questions on them.