

Senior Data Scientist: ([Decimal Point Analytics](#))

Jun 2022 - present

## LLM Finetuning Question Answering task:

- Did supervised finetuning of **Llama 7B, 13B** models on a custom QA dataset to improve answer accuracy.
- Explored partial finetuning, full finetuning, Low Rank Adaptation (**LoRA**), and **Quantized LoRA** techniques.
- Prepared **RAG** pipeline to extract relevant context to answer the given question from a text corpus.
- Used **HuggingFace, VLLM, Ollama**, and Oracle Cloud Infrastructure (**OCI**) to train and deploy APIs.

## Time series forecasting with deep learning:

- Forecasted traffic volume at toll plazas, WPI index, and demand forecasting of retail products.
- Did feature augmentation, trained models like **NBEATS, NHITS, TiDE, TSMixer, SOFTS**. (Deep Neural Nets)
- Made changes in both SOFTS and TiDE architecture to use **dynamic Future Covariates**, improved test MAPE
- Modified data loader and training loop to include a gap between the input series and the forecasted series
- Brought in explainability by getting feature importances using gradients of forecasts wrto inputs (GradCAM)
- Used **PyTorch** for models, **MLFlow** for experiment tracking, **DVC, Jenkins, Docker** for MLOps pipelines.

## RAG with hybrid search for financial reports:

- Built an extraction pipeline that extracts relevant text from the text corpus for a high recall use case.
- Used **TF-IDF, BM25**, dense **vector search**, MMR reranking, and RRF for near-exhaustive context retrieval.
- Achieved >80% recall, >90% precision on context extraction from the corpus (recall is vital for the use case)

## GUI agent:

- Built a GUI agent for automating UI test cases by forking open-source projects like Agent-S and UI-TARS.
- Optimized cost and speed with local SLM (**Gemma3**), without hurting task and action completion rate

## Computer Vision:

- Made an **object tracking** pipeline using **DeepSORT** with **YOLOv8** backend, added object **counting** on top.
- Finetuned Table Structure Recognition (**TSR**) model (**TFLOP**) for a specific table type for an in-house tool.

Worked on multiple agents using RAG, LLMs, DB queries, tried out various SLMs for cost savings for same tasks.

Guided multiple interns and junior data scientists; Recognized by the [Fast Track Career](#) program at DPA.

**Others:** I worked as an educator, mentor at Ed-tech companies, and website editor for the IITK ME website.

Master of Technology	IIT Kanpur	CPI 7.75	2020-22
Bachelor of Technology	IIT Patna	CPI 7.26	2015-19

## **Publication:** [Computer-Vision-Based Technique for Prestress Measurement in Beams](#)

- Designed and demonstrated a cost-effective, non-destructive method to measure prestress in beams.
- This method utilizes a high-speed camera to capture the vibrations of the beam induced by impact loads.
- Digital Image Correlation (**DIC**) and **FFT** are used to gather natural frequency, which predicts prestress.
- These natural frequencies, along with the beam dimensions and material, will predict the prestress
- Published this work in **IEEE Transactions on Instrumentation and Measurement** (Impact Factor: 6)

Implemented **CNN** (digit recognition), worked on building bots for games with RL - **DQN, DDPG, PPO**

**Skills:** Deep Learning, Agentic AI, MCP, MLOps - deployment and monitoring, DSA, OOP, clean and vectorized code, problem-solving, good communication (English, Hindi, Telugu), team player

**Programming Languages:** Python, Java, C, Arduino, MATLAB

**Technologies & Libraries:** NumPy, Pandas, PyTorch, TensorFlow, LangChain, LangGraph, MLFlow, DVC, GitHub, Linux, AWS, OCI, Docker, Jenkins

AIR 208 in GATE ME 2021, and AIR 4102 in JEE Advanced 2015. My hobbies are drones, robots, and 3D printers