

Anudeep Peela

Bangalore, India | anudeeppeela9@gmail.com | +91-93813-42465 | anudeep-peela.github.io |
linkedin.com/in/anudeep-peela | github.com/anudeep-peela

Professional Summary

- Machine Learning Scientist with **3.5+ years** building and deploying production ML, NLP, and GenAI systems in **Python, SQL, PySpark, PyTorch**, and **scikit-learn** across finance, pharma, customer operations, and manufacturing.
- Built retrieval-assisted and query-understanding-adjacent systems using **hybrid search, RAG, transformers**, and **fine-tuned LLMs**; partnered with engineers, analysts, and stakeholders to deliver **+\$27M ARR uplift, \$6.3M/yr savings, 90% faster reviews, and 30% lower call volume**.

Skills Matrix

ML & Statistics: Regression, clustering, XGBoost, neural networks, supervised/unsupervised learning

Search / NLP / GenAI: Hybrid search, RAG, embeddings, semantic retrieval, transformers, intent & sentiment analysis, LangChain, LlamaIndex, OpenAI Agent SDK, LoRA/PEFT fine-tuning

Programming & Data: Python, SQL, PySpark, distributed pipelines, Azure Data Lake

Frameworks & Production: PyTorch, TensorFlow, scikit-learn, MLflow, Docker, Kubernetes, Terraform, Plotly Dash

Cloud: Azure, AWS, GCP

Experience

McKinsey & Company

Bangalore, India

Data Scientist II

May 2024 – Present

- Investigation AI Co-Pilot (Pharma):** Designed a retrieval-assisted NLP system using **hybrid search** over investigation records and LangChain human-in-the-loop workflows to surface relevant historical evidence for root-cause analysis; reduced review time by **90%**, improved output accuracy to **96%**, shortened closure time by **40%**, and cut repeat critical deviations by **30–40%**, saving **\$0.8M/yr**.
- Production Multi-Agent System (Revenue Management):** Built and deployed an agentic cash-reconciliation workflow using **OpenAI Agent SDK** with custom tool calling; productionized on **Azure** with **Docker** and **Terraform**, eliminating manual reconciliation and delivering **\$6.3M/yr** in estimated savings.
- Optimization Pipeline (Manufacturing):** Developed and deployed an end-to-end **XGBoost** pipeline over **2,000+** sensor tags via **PySpark** on **Azure Data Lake**; implemented automated daily retraining and operational monitoring with **MLflow**, reducing import variability by **40 t/h** and yielding **\$2.5M/yr** savings.
- Elasticity-Driven Pricing Optimization (Retail Banking):** Built a production decision system clustering customers by price elasticity to optimize personal and auto-loan pricing; partnered with client leadership to land adoption and drive **+\$27M** in annualized revenue while maintaining retention.

Data Scientist I

Jul 2022 – Apr 2024

- Voice Analytics (NLP / Query Understanding):** Built and deployed a transformer-based NLP pipeline over call-center transcripts for **intent** and sentiment analysis, helping redesign IVR and proactive outreach flows; reduced call volume by **30%** and delivered **\$5M** in savings.
- LLM Fine-Tuning & Code Generation:** Fine-tuned **StarCoder**, **Llama 2**, and **Falcon 40B** with **LoRA/PEFT** on proprietary ABAP datasets; built an evaluation harness with human refinement loops, automating **25–50%** of legacy code generation.
- Optimization Engine (Automotive):** Developed a config-driven optimizer integrated with **CATIA** to automate engine-layout design (parts allocation and piping); reduced design iteration cycles from **2 months** to **<40 min**.

Education & Certifications

Indian Institute of Technology Bombay

Jul 2017 – Jun 2022

B.Tech. & M.Tech. (specialization: Artificial Intelligence) · CGPA 8.0/10

Certifications: AWS Certified Machine Learning – Specialty; MLOps Engineering on AWS