

Deepak Krishna

9844667500 | deepakkrishna2206@gmail.com | LinkedIn | GitHub

EDUCATION

Presidency University

B.Tech in Artificial Intelligence and Machine Learning; GPA: 8.03/10

Bengaluru, India

Expected Aug. 2027

TECHNICAL SKILLS

Languages: Python, Java, C++, JavaScript

AI/ML & Agentic Frameworks: PyTorch, TensorFlow, Transformers (Hugging Face), LangChain, LangGraph, n8n, Scikit-Learn, NumPy, Pandas, MediaPipe, Ollama

Core Competencies: Machine Learning, Deep Learning, Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), Prompt Engineering, Model Quantization (FP32 to INT8), Computer Vision, Debugging

Backend & Databases: FastAPI, RESTful APIs, SQL, MySQL, MongoDB, Qdrant (Vector DB)

Frontend & 3D: React, Three.js

Tools & Practices: Git, GitHub, Docker, AWS, Agile Development

PROJECTS

Sentira-3D | *Real-Time Sentiment Analysis Engine* | Python, PyTorch, FastAPI, React, Three.js

- Developed a high-performance sentiment classification system combining an asynchronous FastAPI backend with an interactive 3D particle visualization frontend.
- Trained and fine-tuned a DistilBERT (66M parameter) transformer model on a 1.6M tweet dataset using mixed precision (FP16), achieving 88.5% accuracy and an ROC AUC of 0.9546.
- Optimized inference latency to <10ms on NVIDIA GPUs via FP32 to INT8 dynamic model quantization and cross-platform hardware support (CUDA, Apple MPS, CPU).

NariConnect | *AI-Powered RAG Matching Engine* | Python, FastAPI, Ollama, Qdrant, MongoDB

- Built a Retrieval-Augmented Generation platform for semantic search over unstructured documents using local LLMs, vector embeddings, and enterprise-scale information retrieval.
- Developed an asynchronous FastAPI backend integrating Qdrant for intelligent vector search and MongoDB for structured program metadata, enabling natural language querying over unstructured data.
- Built a custom data pipeline to scrape, clean, and vectorize complex datasets, enabling automated profile extraction.

Accident Severity Prediction | *End-to-End ML Pipeline for Imbalanced Classification* | Python, Scikit-Learn, Streamlit

- Built an end-to-end machine learning pipeline to classify accident severity (Slight/Serious/Fatal) from ~12,000 records across 30+ categorical features, addressing extreme class imbalance (~1% fatal cases).
- Benchmarked Logistic Regression, Random Forest, XGBoost, LightGBM, and CatBoost; selected LightGBM and achieved a Macro F1-score of 0.506 via custom threshold-weighted probability optimization and rule-based overrides.
- Deployed an interactive Streamlit application for real-time severity prediction with live class-probability visualization.

SignSpeak | *Real-Time Sign Language Translator* | Python, MediaPipe, Scikit-Learn

- Developed a real-time computer vision system to translate sign language into text by extracting hand landmarks using MediaPipe.
- Performed feature engineering on MediaPipe hand landmarks before training and evaluating SVM and Random Forest models, achieving 99% inference accuracy.

AWARDS & HACKATHONS

1st Place, Innovative Projects using Raspberry Pi (IPR): Secured first place out of 506 competing teams.

2nd Place, Datathon (Innovatex 4.0): Runner-up at the International Level Tech Fest hosted by Presidency University.

Hackathon Participant, SheLeads 2.0: Built and deployed the NariConnect RAG platform during the Harvest Club hackathon at Presidency University.

Hackathon Participant, Rift'26: Competed in the National Level Hackathon organized by Physics Wallah.

Hackathon Participant, Build with Trae: Developed a financial management tool focused on budgeting, expense tracking, and financial insights.