

Pradeep Bolleddu

Boston MA, 95013 | +1 774-635-8035 | bolleddu.pradeep@gmail.com | [@linkedin](#) | [@github](#) | [@pradeep.com](#)

WORK EXPERIENCE

- Applied Scientist | Oliv AI, San Francisco, California
- Sep 2024 - Present
- Addressed **45%** False positives in legacy transaction monitoring system. Engineered Python/TensorFlow models (XGBoost, Isolation Forest) with Azure ML(*Blob Storage, Azure Functions & Azure Synapse Analytics.*) pipelines & Deployed Monitoring via **Azure Machine Learning** to Reduce False Positives by **62%** while maintaining **98%** Recall helps Accelerating model Deployment cycle from **6** weeks to **9** days.
 - Solved **70%** unresolved Queries in a **Novel LLM** system requiring cross-domain knowledge. Designed a OpenAI GPT-4 API **RAG System** with Hugging Face embeddings, Amazon Textract-OCR, Pinecone vectorDB, & LangChain orchestration and Achieved **89%** first-contact resolution (**+49%** YoY), cutting average handle time by **40%** (**\$500K** annual savings).
 - Conducted approximately **200** A/B tests and experiments with LLM models, significantly enhancing online user metrics and making the search page the most engaging element for users.
 - Trained churn prediction and price sensitivity models with XGBoost/CatBoost, using hyperparameter tuning and A/B testing for pipeline optimization.

- Machine learning Researcher | Umass Chan Medical School, Dartmouth, MA
- November 2023 - July 2024
- Implemented Image classification techniques on cancer datasets by Analyzing **5K+** merged Brain MRI images through an Cloud based **ETL** pipeline on Google Cloud Platform using **Cloud Dataflow**. Achieved **97%** accuracy which significantly Improved Diagnostic precision for oncology teams.
 - Designed a Bayesian-optimized ensemble (**XGBoost, CNN, Autoencoder**) for cancer prognosis, Achieving a **0.987** F1 score (**18%** fewer false positives) and Enabling personalized treatment plans for **20K+** annual patients.

- Quantitative Data Scientist | The Perfect Future, Delhi, India
- May 2022 - Sep 2022
- Developed **13** Tableau dashboards for Marketing Analytics and **6** Power BI dashboards for Financial reporting which enhanced CRMs strategies by **20%** and collaborating with sales to boost conversion rates by **15%**.
 - Automated Database Operations Using **Azure SQL Database** by increasing operational efficiency by **20%** through **170+** CRUD operations, **20** database triggers, and **10** stored procedures for data management & delivered actionable insights through **Looker Studio** for Reporting.
 - Designed and trained a user classification model specifically for ranking systems within the A/B testing framework, leading to a **40%** increase in test sensitivity and a more accurate evaluation of ranking search model performance.

EDUCATION

University of Massachusetts	Master's in Artificial Intelligence & Data Science	August 2023 - August 2024
Indian Institute of Technology- IIT Ropar	BTech in Engineering & Minors in Computer Science	July 2019 - May 2023

TECHNICAL SKILLS

Programming Languages : Python, R, SQL (SQL Server, Oracle SQL & T-SQL.) & Scala.
Frameworks : Tensorflow, Scikit-learn,PyTorch, Keras, JAX,Hugging Face Transformers & OpenCLIP.
DevOps and Security Tools Git, Docker, OpenShift, Kubernetes,Eclipse & Linux (Configuring and Managing).

PROJECTS

- Autonomus Vechile Wardbot for Covid Patients
- Patent
- Leveraged Deep Reinforcement learning (DQN) and YOLOv5-based computer vision detection to develop **Wardbot**, delivering food, medicine, and equipment to over **200+** COVID 19 isolation ward patients. [Slides & Report](#)
- Multimodal Emotion Recognition Transformer
- Github code
- Engineered a Multimodal Transformer fusing [Video frames](#) , [Speech Audio](#) and Text transcripts for real-time emotion classification, Achieving a **12%** Improvement in weighted F1-score on **IEMOCAP** by Implementing hierarchical cross-modal Attention and semantic role-aware graph reasoning.
- Optimized Large-Scale LLM Training: DeepSpeed & HuggingFace Accelerate Integration
- Github Repo
- Built** A distributed multi-GPU training pipeline with DeepSpeed ZeRO and HuggingFace Accelerate, tripling throughput and cutting end-to-end training time by **40%**.

RESEARCH PUBLICATIONS

- Authorized [An Analysis of Clustering Techniques](#) leveraging clustering algorithms on **10M+ financial records** to derive Impactful insights for Customer Segmentation Data.
- Architected [Hadoop: An Overview of Data Security](#),Designed **Hadoop Architecture** MapReduce-based solution focusing on **Data Integrity and Privacy**.