ABHIJEETH KOLLARAPU

abhijeethkollarapu@gmail.com | (813) 452-8290 | https://www.linkedin.com/in/abhijeeth-kollarapu | https://github.com/Abhijeeth8

Education

University of South Florida, Tampa, Florida **Aug 2023 - May 2025**

Master of Science in Computer Science (**GPA: 3.83 / 4.0**)
Related Coursework: Machine Learning, Deep Learning, NLP, OS, Computer Architecture, Algorithms, Social Media Mining.

Matrusri Engineering College, Hyderabad, India **Jul 2018 - Jun 2022**

Bachelor of Engineering in Computer Science and Engineering (**GPA: 3.5 / 4.0)**
Related Coursework: Data structures using C++, Java, Design and Analysis of Algorithms, Computer Networks

Experience

Vistex Asia Pacific Pvt. Ltd. - Associate Developer (Hyderabad, India) **Jul 2022 - Jun 2023**

- Worked on an end-to-end topic modeling project involving preprocessing of textual data using NLTK and Spacy and application of DL algorithms to classify support tickets to respective departments. Improved model performance by 1.6% through application of Latent Dirichlet Allocation (LDA), feature engineering, and hyperparameter tuning using optuna.

Vistex Asia Pacific Pvt. Ltd. - Associate Developer Intern (Hyderabad, India) **Apr 2022 - May 2022**

- Trained in NLP workflows including preprocessing, model building, and evaluation using ANN and RNNs.
- Shadowed senior engineers on real projects to learn industry standards and gradually contributed to development tasks by assisting in testing and validating models under the supervision of senior engineers.

COIGN Consultants Pvt. Ltd. - Machine Learning Intern (Hyderabad, India) **Jan 2022 - Feb 2022**

- Developed a **Density-Based Smart Traffic Control System** utilizing Canny Edge Detection to gather traffic data and estimate optimal green signal durations.

HPCL Visakh Refinery - Machine Learning Intern (Hyderabad, India) **Dec 2021 - Jan 2022**

- Optimized ML models (Random Forest, Logistic Regression, and XGBoost) through in-depth EDA, data preprocessing, feature selection and extraction, and hyperparameter tuning using GridSearchCV, achieving 85% accuracy while reducing training time.

Projects

Knowledge Distillation in NLP: Transferred 97% of knowledge from BERT to DistilBERT and from GPT-2 to DistilGPT using advanced distillation methods across four classification tasks using pytorch, huggingface and CUDA.
Large Language Model (LLM) Optimization: Fine-tuned LLMs using LoRA, quantization, gradient accumulation, and checkpointing, achieving more than 50% memory optimization.
Large Language Model (LLM) Applications using LangChain and LangGraph: Built LangChain-based ReAct agents and Retrieval-Augmented Generation (RAG) systems using Chroma/FAISS/Pinecone vector stores.
Automated essay writing and tweet generation via iterative critique using LangGraph.
Built advanced Retrieval-Augmented Generation (RAG) systems that uses multi-query strategy for retrieving relevant documents from Pinecone vector store using LangGraph, which can also browse the internet for real-time, context-aware content.
Text Classification: Developed ML/DL models (XGBoost, GB, SVM, CNN, LSTM) to perform Twitter disaster detection, achieving up to 98% accuracy with feature engineering.
Image Classification: Achieved 89.5% accuracy in brain tumor classification using VGG19 on MRI images.
MLOps & Model Tracking: End-to-end ML tracking on diabetes classification with MLflow (local and AWS), including experiment tracking, project packaging, artifact creation and storage, model and data versioning, and drift detection.
Cloud Deployment (AWS SageMaker): Preprocessed data with AWS Data Wrangler and processors and deployed XGBoost and TensorFlow estimators via built-in and custom Docker images for real-time and batch inference. Built automated ML pipelines using SageMaker's processor, estimator, and evaluator for end-to-end workflows.

Certifications

- AWS Certified Cloud Practitioner (CLF-02)

Technical Skills

- **Programming and Development**: Python, Java, JavaScript, Data Structures & Algorithms, FastAPI
- **Statistics**: Probability Distributions (PDF and PMF), Hypothesis testing, Confidence Intervals
- **Data Engineering and Analysis**: HDFS, Hive, Apache Spark, Databricks, Azure Data Factory, Azure Data Lakes, PowerBI, Tableau
- **Machine Learning**: Supervised, Unsupervised, Reinforcement Learning, Scikit-learn, GridSearch, Hyperopt, Exploratory Data Analysis (EDA), Feature Engineering, NumPy, Pandas, Matplotlib, Seaborn, Plotly
- **Deep Learning**: Tensorflow, Keras, PyTorch, ANNs, CNNs, RNNs, Encoder-Decoder Models, Transformers, GANs, Functional API, Keras Tuner, Optuna
- **MLOps**: DVC, MLflow, Dagshub, Amazon SageMaker, Prometheus, Grafana, WhyLogs/WhyLabs
- **Natural Language Processing (NLP)**: Embeddings, NLTK, Spacy, Topic Modeling, Text Generation and Classification
- **Large Language Models (LLMs) & Generative AI**: LangChain, LangGraph, Tools & AI Agents (OpenAI Agents SDK, MCP), Vector Stores, Retrieval Augmented Generation (RAG), Hugging Face, LoRA & Efficient Fine-Tuning, Prompt Engineering
- **DevOps, Automation & Containerization**: Git, Ansible, Jenkins, GitHub Actions, Terraform, AWS, Docker, Kubernetes