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
– *Master's of Science in Computer Science | GPA: 3.83 / 4.0 | Aug 2023 - May 2025*

Matrusri Engineering College, Hyderabad, India
– *Bachelor of Engineering in Computer Science and Engineering | Jul 2018 - Jun 2022*

EXPERIENCE

Vistex Asia Pacific Pvt. Ltd. | Hyderabad, India | *Associate Developer | Jul 2022 - Jun 2023*
– Executed an end-to-end topic modeling pipeline on support ticket data, including preprocessing, cleaning, and transformation using SpaCy.
– Applied Latent Dirichlet Allocation (LDA) to uncover dominant topics per department, achieving a coherence score of 95%.
– Converted unsupervised topic outputs into labeled data by mapping tickets to their most probable topics and engineering keyword-based features.
– Designed a non-sequential neural network with an LSTM branch for ticket text and a dense layer for topic features, improving classification accuracy by 7%.

Vistex Asia Pacific Pvt. Ltd. | Hyderabad, India | *Associate Developer Intern | Apr 2022 - May 2022*
– Gained hands-on experience designing and implementing ML workflows, including data preprocessing, model training, and evaluation using supervised and unsupervised techniques such as ANNs and RNNs.
– Focused on NLP tasks within IT service data, contributing to real-world classification problems like support ticket categorization and issue prioritization.
– Developed a baseline KNN classifier using structured ticket metadata and extracted features, and assisted in model testing and optimization under senior guidance, achieving minor accuracy gains (< 2%)

COIGN Consultants Pvt. Ltd. | Hyderabad, India | *Machine Learning Engineer Intern | Jan 2022 - Feb 2022*
– Designed and prototyped Smart Traffic Control System using Canny Edge Detection to estimate real-time vehicle density.
– Computed optimal green signal durations based on density estimates, achieving up to 90% accuracy in vehicle detection during simulation.

HPCL Visakh Refinery | Hyderabad, India | *Machine Learning Intern | 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 | *PyTorch*, *HuggingFace, Knowledge Distillation, NLP*
– Transferred 97% of knowledge from BERT to DistilBERT and from GPT-2 to DistilGPT using PyTorch, Hugging Face, and CUDA, across four diverse classification tasks to evaluate which distillation techniques are most effective for each task.
– Performed advanced distillation techniques such as attention weight distillation, intermediate layer distillation, and logits distillation to efficiently transfer knowledge from various components of the teacher model, resulting in an accuracy improvement of nearly 8%.

ArXiv Research papers Chatbot MCP Server with RAG Architecture | *MCP, Langchain, RAG, AgenticAI*
– Built a research paper chatbot and summarizer using RAG (LangChain and Chroma vector store) within a custom MCP server that supports paper embedding and contextual retrieval for 1,000+ academic research papers, reducing researchers' literature review time by up to 40%.
– Integrated the server with Anthropic Claude, enabling tool usage from Claude's interface for real-time document interaction.

Autonomous AI Developer Crew using CrewAI | *CrewAI, AI Agents, Python*
– Architected a multi-agent system using CrewAI where an Engineering Lead agent interprets natural language requirements to orchestrate Python developer agents that generate modular backends and Gradio frontends with full design specifications.
– Integrated CrewAI's built-in and custom tools to write code to the local directory and auto-launch the Gradio UI, enabling one-click deployment of fully functional applications accelerating end-to-end application delivery and cutting development effort by over 90%..

Quora duplicate question pairs classifier | *SKLearn, Tensorflow, NLP*
– Contructed baseline classifiers using XGBoost and Random Forest, encoding texts with Bag of Words, and achieved 71% accuracy.
– Implemented advanced deep learning strategies and achieved almost 80% validation accuracy:
  – First approach generated independent contextual embeddings via attention layers and compared them using cosine similarity.
  – Second approach combined both questions, generated a joint contextual embedding, and classified using a fully connected layer.

Alter-Ego Agent: Self-Representing Chatbot with Email Drafting | *AI Agents, RAG, APIs, Langchain*
– Created a personalized chatbot using LangChain that impersonates me to answer questions about myself leveraging RAG technique.
– Created a Gmail-integrated tool that is invoked by an email-specialist LLM agent to send a professional connection email if the interviewer is interested and shares their email, facilitating timely and effective networking while reducing manual effort by 90%.

Topic Modeling Resumes to Predict Applicant Job Intent | *Unsupervised Learning, LDA, SKLearn*
– Employed Latent Dirichlet Allocation (LDA) for unsupervised learning to classify resumes based on inferred job intent achieving a topic coherence of 93%.
– Identified key recurring terms in resumes targeting similar job roles and approximated the applicant's job preference accordingly.

Large Language Model (LLM) Applications using LangChain and LangGraph | *LangChain, LangGraph, Tools*
– Engineered LangChain-based ReAct agents and built automated essay writing and tweet generation applications through iterative critique mechanisms using LangGraph, decreased the workload by 95%, resulting in substantial time savings.
– Designed an advanced RAG system with multi-query retrieval strategy using LangGraph and Pinecone, boosting retrieval effectiveness by 25% on unclear queries, capable of browsing the web to fetch real-time, context-aware information, improving generation accuracy.

Text Classification for Disaster Detection | *SKLearn, TensorFlow, NLP*
– Developed multiple ML and DL models, including XGBoost, SVM, CNN, and LSTM, to classify tweets related to disasters.
– Applied advanced NLP techniques and feature engineering (TF-IDF, word embeddings, POS tagging) to extract relevant patterns from noisy Twitter data and achieved up to 98% classification accuracy, optimizing model performance through hyperparameter tuning and ensemble approaches.

Image Classification of Brain Tumors | *VGG19, CNN, Medical Imaging*
– Designed an end-to-end classification pipeline to detect brain tumors from MRI scans using a fine-tuned VGG19 architecture.

- Conducted preprocessing using OpenCV (e.g., grayscale normalization, contrast adjustment) and applied data augmentation to improve model generalization and achieved 89.5% accuracy, outperforming baseline CNN models and reducing false positives in tumor detection.

MLOps and Model Tracking Pipeline | *MLflow, AWS, Scikit-learn*
- Built a complete ML tracking system for diabetes classification using MLflow on both local and AWS environments.
- Tracked experiments, parameters, and metrics; managed model and dataset versioning; and monitored drift for ongoing model quality assurance.
- Packaged the project with Docker and set up reproducible training environments for collaboration and deployment.

Cloud Deployment on AWS SageMaker | SageMaker, Data Wrangler, XGBoost, TensorFlow, Docker
- Preprocessed diabates dataset with AWS Data Wrangler and SageMaker Processing for cleaning and feature transformation.
- Deployed XGBoost and TensorFlow models via SageMaker built-in containers and custom Docker images for real-time and batch inference.
- Automated the ML lifecycle using SageMaker Pipelines with processors, estimators, and evaluators to enable reproducible, end-to-end deployment workflows.

CERTIFICATIONS

AWS Certified Cloud Practitioner (CLF-02)

TECHNICAL SKILLS

Programming and Development: Python, Java, JavaScript, Data Structures & Algorithms, SQL, FastAPI

Data Engineering and Analysis: Hadoop, Apache Spark, Databricks, Azure Data Factory, PowerBI, Tableau

Statistics: Probability Distributions, Statistical Inference (Confidence Intervals and Hypothesis Testing), Regression Analysis

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, Neural Networks, Transformers, GANs, Functional API, Optuna, OpenCV

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

Cloud & DevOps: Git, Ansible, Jenkins, GitHub Actions, Terraform, AWS, Docker, Kubernetes