Prathyush Kumar Reddy Lebaku

+1 713-873-1975 | Houston,TX | plebaku@cougarnet.uh.edu | linkedin | github | Portfolio

Summary

Data Scientist — Data Engineer — AI Engineer — Data Analyst
Worked as a Data Analyst and AI engineer with a strong understanding of deep learning models and large language models (LLMs), with a keen interest in data extraction, preprocessing, and deploying intelligent solutions for real-world applications.

Education

University of Houston Houston, TX
Master of Science in Engineering Data Science (CGPA: 3.89/4.0) *Aug 2023 – May 2025*
Coursework: Deep Learning, Data Structures and Algorithms, Advanced Machine Learning, Probability and Statistics, Computer Vision, Natural Language Processing, Big Data Processing, Artificial Intelligence, Data Analysis, OOP(Object-oriented programming).

Vellore Institute of Technology Vellore, India
Bachelor of Technology in Computer Science and Engineering (Specialization in Data Science) *June 2019 – May 2023*

Experience

Data Analyst Research Assistant Sep 2023 – Jan 2024
Texas Department of Transportation(TxDOT),USA affiliated with University of Houston *Houston, TX*

- Conducted in-depth analysis of pavement conditions and traffic crash characteristics using advanced statistics. Used BeautifulSoup for coordinate extraction and GeoPandas for interactive crash location maps.
- Developed negative binomial and ordered probit models to predict crash frequency and severity, aiding Iowa State in securing $200,000 for road infrastructure improvements. Research Paper Link GitHub

Machine Learning/AI Research Assistant Jan 2024 – Present
University of Houston *Houston, TX*

- Optimized CRDDC2024 dataset using fine-tuned Faster R-CNN and YOLO (v5x–v10x), achieving 70% F1 score overall and 76% on USA dataset.Additionally, hyperparameter tuning was applied. Git hub
- Developed a Visual Question Answering(VQA) system using LLaMA-1.5v-13b(LLM), YOLO World for zero-shot object detection, and a depth estimator(MiDaS), to answer questions about Objects within 15 feet. Git hub
- **Currently working** on classifying environmental conditions. using three different approaches CNN, Vision Language Models with BERT pipeline and direct Vision Language Model to classify .

Personal Projects

Low Resolution Pavement Condition Classification(CNN) | *Transfer Learning,Ensemble Learning* May 2024 - Jul 2024

- Improved satellite image classification accuracy by 11% (from 80% to 91%) using transfer learning on 12 CNN models compared to traditional fine-tuning techniques. Applied ensemble learning with the top 4 models (MobileNetV2, ConvNextBase, ResNet50, InceptionV3) to classify 5 crack types.
  Conference Presentation: Presented findings at AISIM-2024, University of Colorado, Boulder, USA.Research Paper , Git Hub

Real Time Action Recognition using Human Key points(LSTM) | *Mediapipe,OpenCV,TensorFlow* Aug 2024 - Sep 2024

- Collected and processed 90 balanced videos by extracting 1662 keypoints (hands, face, pose) per frame using Mediapipe over 30 frames/video. Stored data in NumPy arrays for a lightweight, efficient alternative to CNN-based models.
- Developed a real-time ASL recognition system with 98.5% F1 score using a stacked LSTM model (3 LSTM, 3 dense layers) with dropout and regularization to classify 3 gestures .

Context-Aware Question Answering System Using LLMs | *Hugging Face,Flask,Colab* May 2024 – Jul 2024

- Fine-tuned **BERT Base** and **BiDAF** models on the **SQuAD v2.0** dataset for building a QA system.
- BERT model outperformed BiDAF by leveraging its **transformer-based architecture**, achieving an **Exact Match (EM) score of 60%** and an **F1 score of 64%**. .
- Deployed the fine-tuned BERT model as an interactive web application using **Flask**, enabling real-time question-answering.

Multi-Agentic Retrieval-Augmented Generation (RAG) System | *AutoGen, HuggingFace, ChromaDB, Streamlit*

- Engineered a multi-agent RAG system using **AutoGen**, coordinating agents for retrieval, context analysis, generation, evaluation, and query refinement.
- Used **HuggingFace embeddings** and **ChromaDB** for similarity-based context retrieval from document datasets.
- Designed a dynamic state_transition mechanism for adaptive agent workflows based on evaluation feedback.
- Built an interactive UI using **Streamlit**, enabling real-time query answering with refinement loops and response scoring.

Technical Skills

Programming Languages Python, SQL
Data Extraction & ETL: PySpark, Apache Airflow, AWS Glue
Data Preprocessing & Visualization: NumPy, Pandas,OpenCV,Pillow, spaCy, NLTK, Seaborn, Matplotlib, Tableau
Machine Learning & Deep Learning: Scikit-learn,Ultralytics, TensorFlow (with Keras), PyTorch, Hugging Face Transformers
Cloud/Deployement Tools: AWS (Glue, Redshift, SageMaker,EMR),Flask, Google Colab, Git/GitHub,Streamlit

Certifications

AWS Data Engineer: 1NHEHYJKOINV32, **Advanced Deep Learning(Andrew NG)**, **Advanced SQL**