HAIDER ALI

Hali4@mail.yu.edu • https://www.linkedin.com/in/haiderzali/ • https://github.com/AliHaider20
Austin, TX 78703 • (551) 344-4769

EDUCATION

Yeshiva University Master of Science, Artificial Intelligence **May 2024**
Coursework: Numerical Analysis, Computer Vision, Statistics and Probability, Machine Learning, Generative AI

EXPERIENCE

S&P Global – *Machine Learning Joint Research Project;* NY, NY **May 2023 – August 2023**
- Scraped, engineered, and extracted relevant features of Global Precipitation Data from NASA's GDDP (HDF) files using R
- Employed Morphological image analysis, and applied KNN and CNN models to extract features
- Evaluated multiple super-resolution models; **ESRGAN** maintained 80% image quality while enhancing resolution
- Utilized **AWS** for **MLOps** orchestration to deploy the model, focusing on scalability and maintainability, and enabling ongoing monitoring through integrated **MLFlow**

Kan Innovations – *Machine Learning Engineer;* Mumbai, India **January 2022 – July 2022**
- Crafted a large image dataset by scraping foot images, merging pressure mat data, and employed Pix2Pix GAN augmentation
- Leveraged **DenseNet121**, **DeepLabV3+**, and custom segmentation models to estimate the pressure points in a foot
- Constructed 3D voxel foot meshes using **Open3D** library and Blender to help design insoles, reduced **40% foot** deformity
- Implemented **BlazePose** model for real-time pose estimation, enhancing gaming interaction reducing **20%** foot deformity

PROJECTS

University Chatbot using LLM ***[github]*** | ***Large Language Models*** | ***Data Engineering*** **September 2023**
- Scraped, augmented, and created 60000 question-answer pairs manually and using the **NLPAug** tool, **BERT**, and **GPT** models
- Fine-tuned and merged including Microsoft's **Phi 1.5B**, **Flan T5**, **GPT 2**, and **Mistral Instruct** achieving 70% RougeL
- Compressed Mistral Instruct model with DeepSparse for faster inference, applied Prompt Engineering for optimal responses
- Deployed GPT2 model on local server with RAG pipeline using **LlamaIndex** and **FAISS**

Cow TEAT Keratosis Level Identification ***[github]*** | ***Data Augmentation*** | ***Generative AI*** **March 2023**
- Employed SURF, SIFT, HOG, and AutoEncoders for image feature extraction, resulting in a **50% F1 score** with SVC and KNN
- Balanced data using **Stable Diffusion**, utilizing SOTA models including **ResNet 50** as a reference point for comparison to achieve an **80% F1 score** using GoogleNet

Medical Expense Prediction ***[github]*** | ***Data Analysis*** | ***Healthcare*** **November 2022**
- Analyzed correlation between non-smoking individuals ages 0-19 and 7% lower medical charges, mostly non-smokers
- Identified a significant association between high BMI (>30) and smoking, resulting in increased medical expenses
- Utilized Linear Regression with forward selection to extract key features, yielding a **90% R2 score** using **XG Boost**

ACHIEVEMENTS

- A top performer in consecutive Python Hackathon events with over 500 participants
- Gold Medalist in GreyAtom's Data Science Competition among 100 participants

TECHNICAL SKILLS

- **Programming and Tools:** Python, R, AWS, Matplotlib, SQL, RAG, NLP, C/C++, Spark, PyTorch, Airflow, TensorRT, MLflow, Docker, Git, ONNX, LLamaIndex, DevOps, Langchain, FAISS, Optimization, CI/CD, RESTful API
- **Data Science Techniques:** OCR, Prompt Engineering, Machine Learning, Deep Learning, NLP, Computer Vision, MLOps
- **Machine Learning:** XGBoost, SVM, Decision Forests, LightGBM, PCA, t-SNE, DBSCAN, Statsmodels API, LSTM, VAE, VIT, Random Forests, AdaBoost, Generative AI, LLM, BERT, RoBERTa, GPT, Stable Diffusion, GAN