Hema Varshita Muppalla

hvarshita.github.io hvarshita hema-varshita-m hmuppa@uw.edu 206-387-9631

Education

- **University of Washington Seattle** Seattle, WA
  MS in Data Science ; GPA: 3.96/4.0 *Sep. 2022 – Mar. 2024*
  Courses: Machine Learning, Experimental Design, Deep Learning, Computer Vision, Data Visualization
- **Indian Institute of Technology Hyderabad** Hyderabad, India
  BTech in Electrical Engineering with a Minor in Economics; GPA: 8.21/10.0 *Aug. 2017 – May. 2021*
  Courses : Data Structures, DBMS, Econometrics and Forecasting, Convex Optimization, Image Processing, Data Science Analysis

  – Among the top 150 students across India to attend the **Google Research Summer School (2020).**

Professional Experience

- **Amazon Web Services - AI** Santa Clara, CA
  Software Engineer (Machine Learning) Intern *Jun 2023 - Sep 2023*
  - **Recommender Systems**: Implemented Serverless Model Training Workflows for Recommendation Algorithms for AWS Personalize. Leveraged Infrastructure as Code principles to optimize infrastructure deployment.
  - **Workflows**: Migrated the workflows to AWS Step Functions & introduced support for **parallel model training**, mitigating two-phase deployments, addressing scaling issues, visualizing workflow components & cutting debugging time by **50%**
- **Publicis Sapient** Bengaluru, India
  Senior Data Scientist *Aug 2021 - Sep 2022*
  - **Aspect Based Sentiment Analysis**: Scraped employee reviews from anonymous review platforms and labelled over 1500 data points. Trained T5 and DistilBERT models and experimented with merging with benchmark datasets achieving best f1-score of *0.80 & 0.73* for negative & positive sentiments respectively.
  - **Conversational AI**: Built the cognitive intelligence chatbot platform using *Elasticsearch, Rasa and AWS* to answer queries from user uploaded / client-specific knowledge base. Deployed to **FastAPI endpoint** for faster retrieval. The cognitive platform used for customer service resulted in **1.84x** more conversions.
  - **Database Management**: Encrypted passwords on MongoDB backend using salting mechanism to secure **thousands of user accounts** across two platforms. Implemented a robust authorization and authentication system using row-level locking.
  - **Customer Analytics**: Leveraged query results from Azure SQL to perform preliminary data visualizations and conducted advanced analyses, such as Anomaly, Outlier Detection, and Clustering Analysis, on demographic and behavioral attributes to identify potential customers resulting in **23%** reduced marketing costs.

  Data Science Intern *Jun 2020 - Jul 2020*
  - **Retail Video Analytics**: Built an end to end deep learning pipeline using *3D ConvNets*, *Keras OpenPose* and a *transfer learning* model built using a *VGG based network* to generate **real time actionable business insights** from retail store CCTV footages. The model detects metrics - footfall & product interaction count with an *accuracy of 99.9%*

Projects

- **Self-Supervised Learning for Visual Data**: Developed a self supervised learning framework using SimCLR for object classification and detection using subsets of ImageNet dataset outperforming baseline supervised classification algorithm by 5%.
- **Researching Recession Proof Retail Trades**: Conducted research on the impact of recessions on retail trades and analyzed sales trends during economic fluctuations. Applied statistical methods, such as Welch test, Linear Regression, and ANOVA to provide data-driven insights into the changes in retail trades during recessions.
- **Visual Question Answering** : Integrated the VGG16 network for image understanding and an LSTM model for textual understanding on subset of VQA dataset to generate probability distribution of answers with a top-1 accuracy of 43%.
- **Machine Learning Algorithms**: Implemented K-means clustering on images, Principal Component Analysis for decorrelation, Maximum Likelihood Estimates, Gaussian Mixture Models, and Sparse Auto Encoder *from scratch* (using only Numpy, Scipy).
- **Fine-Tuning Large Language Models (LLMs)**: Fine-tuned a Dolly LLM on an FAQ dataset containing instruction and response fields using LoRA for efficient adaptation and evaluated the performance using metrics: cross entropy and perplexity. Analysed the LLM for bias and hallucination and deployed the fine-tuned model using Amazon Sagemaker.
- **ML based User Localisation and Activity Recognition**: Built a ML pipeline combining Random Forests and Neural Network architectures to predict a smartphone user's **indoor location** and activity performed based on WiFi Signal Strength and Sensor Data collected from Gyroscopes and Accelerometers with *90% and 96% accuracies* respectively to identify anomalous activities.

Skills

- **Domain**: Machine Learning, Data Science, Vision & Image, Language Processing, MLOps
- **Languages & Web**: Python, C/C++, Java, TypeScript, R, SQL, Flask, RestAPI
- **ML Frameworks**: Pytorch, Tensorflow, Theano, Caffe
- **Dev** : Elasticsearch, Rasa, Dialogflow, Hugging Face, Docker, AWS S3, AWS Lambda, AWS Step Functions, Azure