DUSHYANT MAHAJAN

dushyantinboston@gmail.com | LinkedIn | Portfolio | GitHub | Boston, MA

Professional Experience

Raga AI **Fremont, CA**

Data Scientist *Jan 2024 - Jul 2024*

- Co-developed the **open-source** framework Raga LLM Hub, enriching it with over 50 comprehensive metrics and establish critical **guardrails** for **LLMs** and **Retrieval-Augmented Generation** (RAG) applications, enhancing model respons**e** accuracy
- Developed an **observability** tool RagaAI Catalyst to provide **trace** recording inside RAG applications, offering a one-click deployable solution that allows for **fine-tuning** and **evaluation** of LLM applications, streamlining deployment processes
- Engineered "**RAG Builder**", a tool with drag-and-drop functionality that enables customizable RAG components construction and optimization for specific use cases, reducing development time for custom **RAG pipelines** by **50%**
- Benchmarked and optimized custom RAG pipelines for **prompt response quality** across **Llama**, **Gemma**, and **Mistral** models, significantly reducing **token costs** while enabling engineering teams to identify the most cost-effective solutions for deployment.

Raga AI **Bangalore, India**

Data Science Consultant *May 2022 – Aug 2022*

- Designed and deployed an API pipeline with dashboard for interactive **visualization** and **clustering** of **DNN** embeddings using techniques like **t-SNE, UMAP** and **PCA**, enabling real-time analysis and interpretation of high-dimensional data
- Implemented Maximum Mean Discrepancy (**MMD**) and **Kolmogorov-Smirnov** tests for **drift detection** in image datasets, reducing undetected **data drift** and enhancing model stability
- Leveraged **AE**, **VAE**, Variational Auto-Encoding Gaussian Mixture Model (**VAEGMM)** algorithms to identify outliers in high-dimensional datasets, improving **anomaly detection** accuracy by **40%**.

Askim Technologies **Mumbai, India**

Software Engineer *Jan 2021 – May 2022*

- Led the development of a multimedia prescription platform, creating a **python pipeline** that processed **high volumes** of paper prescriptions, significantly improving clarity, and reducing patient follow-up queries by **30%**
- Built and deployed a resilient, **full-stack** application on **AWS** using the **MERN** stack, optimized for **scalability** and **security** with **multi-AZ** architecture, **HTTPS**-enabled **CRUD** endpoints
- Crafted a robust **CI/CD** workflow using **GitHub Actions**, integrating **HashiCorp Packer** to automate the creation of latest Ubuntu-based **Amazon Machine Image (AMI)** for web applications, thereby facilitating continuous integration
- Automated the provisioning of **AWS** services - **Route53, VPC, EC2, RDS, S3, SNS, Lambda, DynamoDB, IAM, CloudWatch** with **Pulumi** IaC.

Projects

Parallelization Techniques in Deep Learning for Image Classification using PyTorch | Link

- Leveraged custom CNN architecture and PyTorch to process weather images for classification, achieving **78%** F1-score
- Utilized PyTorch's **DataLoader** and **Dask** for data processing, achieving a **4x speedup** compared to single threaded data-loading
- Employed PyTorch **DP (Data Parallel)** and **DDP (Distributed Data Parallel)** for model training across **1- 4 GPUs** on a HPC cluster, recording a **1.41x speedup**.

Technical Skills

Programming Languages: Python, Go, SQL, JS, HTML/CSS
Cloud & DevOps: AWS, Google Cloud, Azure, Docker, Pulumi, GitHub Actions, Packer
Databases: MySQL, PostgreSQL, MongoDB, DynamoDB, ChromaDB, Atlas Vector Search, Spark
Frameworks and Technologies: Langchain, LlamaIndex, OpenAI API, Ollama, OpenLLMetry, PyTorch, Tensorflow, Transformers

Publication

Dushyant M. et al. (2024). Roux-lette at Discharge Me! Reducing **EHR** (Electronic Health Record) Chart Burden with a Simple, Scalable, Clinician-Driven **AI** Approach. **23rd Workshop on Biomedical Natural Language Processing**, pages 719 - 723, Bangkok, Thailand. **ACL** (Association for Computational Linguistics). https://aclanthology.org/2024.bionlp-1.63/

Education

Northeastern University **Boston, MA**

Master of Science, Information Systems | ***GPA**: 3.63/4.00* *Sep 2022 – Dec 2024*

University of Mumbai **Mumbai, India**

Bachelor of Technology, Computer Science and Engineering *Aug 2015 – Oct 2020*