Vivek Aryan Ravula

Houston, Texas | 3463038568 | aryanravula@gmail.com | LinkedIn | Portfolio

Education

Master of Science in Data Science | GPA: 3.93 — **Houston, TX**
University of Houston, Cullen School of Engineering — *Jan 2023 – May 2024*

Bachelor of Technology in Civil Engineering — **Manipal, India**
Manipal Institute of Technology — *Aug 2015 – May 2019*

Work Experience

AIceberg — **New York City (Remote), USA**
Artificial Intelligence/Machine Learning Research Assistant (Generative AI-NLP) — *June 2023 – May 2024*

- Collaborated with an 8-member interdisciplinary team and engaged with 4 stakeholders to assess relevance and specificity of LLMs.
- Designed production-ready code solutions for AI pipeline and model deployment, seamlessly integrating it with Airflow and AWS S3.
- Fine-tuned large language models (llama 2, mistral, Falcon, GPT) on a custom dataset using QLoRA fine-tuning method.
- Grounded the finetuned LLM with external knowledge from semantic Retrieval Augmented Generation and Knowledge Graphs to generate quality response through prompt engineering to tackle hallucination in LLM.
- Researched and executed experiments across multiple frameworks (LangChain, Llamaindex), SOTA embedder models, and vector storage databases (simple methods to FAISS) to optimize both the speed and quality of processes.

Meesho — **Bangalore, India**
Business Analyst — *Oct 2022 – Jan 2023*

- Implemented A/B testing methodologies to measure the performance of webinars and personalized training sessions, optimizing customer engagement strategies based on empirical insights.
- Provided business recommendations to improve supplier engagement.
- Designed and maintained analytical dashboards on Presto for tracking the L0 metrics of the Supplier Activation charter.

Swiggy — **Bangalore, India**
Business Intelligence Engineer (Product Analytics, Data Engineering) — *Apr 2021 - Aug 2022*

- Collaborated with a 5-member cross-functional vertical, served as the Single Point of Contact (SPOC) for my division, and coordinated with 5 stakeholders towards the success of Customer Intelligence Chatbot products.
- Designed experiment design to conduct advanced analytics techniques, including A/B testing and statistical inference, to assess in-house product features and provided actionable business recommendations.
- Improved the CPO by 15% by changing the nomenclature of a bot disposition. Reduced 95th percentile customer wait times during peak hours by 60% by balancing the load.
- Developed and managed analytical dashboards in Power BI to monitor KPIs, track trends, assess company initiatives, user behavior, and agent performance.
- Conducted in-depth driver analysis to pinpoint bottlenecks opportunities in Swiggy Chatbot flow.
- Conceptualized and formulated the enhancement of various metrics including active agents and bot efficacy metric.
- Spearheaded root cause analysis initiatives to identify underlying factors contributing to reduced notification engagement in customers.
- Collaborated with enterprise data warehouse, data governance, and Cross-departmental teams to tune data pipelines for performance and scalability using ETL processes, Databricks, and Airflow.

Technical Projects

Movie Recommender System using LLMs – *ETL | Recommendation Algorithms | Vector Database | LLM Quantization*

- Designed a movie recommendation system using semantic similarity search and Phi-3. Data was scraped from Wikipedia and TMDB.
- Employed and assessed two methods of producing embeddings, Count Vectorization and embeddings generated by sentence transformer.
- Leveraged LangChain to store embeddings in Weaviate Vector Database and create a pipeline with Phi-3 for generating summaries.
- Developed the front-end and back-end using Next.js, Tailwind CSS, React, and FastAPI.

Chest Cancer Classification deployed on AWS – *Deep Learning | MLOps | CI/CD*

- Developed an end-to-end web application for detecting chest cancer in images, integrating ML lifecycle management tools such as MLflow for experiment tracking and DVC for pipeline tracking and version control.
- Deployed the app on an AWS EC2 instance for scalability and accessibility.

Reddit Sentiment Analysis - *Data Engineering Pipeline*

- Created a streamlined pipeline with Apache Airflow and AWS services to extract, transform, and load Reddit data into Amazon Redshift.
- Conducted Sentiment Analysis on the data, contrasting the effectiveness of rule-based and transformer approaches.

Skills

Programming:	Python, C++ | PyTorch, TensorFlow, Keras, Spark | HTML, CSS, JavaScript (basics)
Database:	SQL**,** MySQL, Snowflake | Vector Database (FAISS, Weaviate) | Graph Database (Neo4j)
Analytical Tools:	Statistics**,** Excel (VBA) | Tableau (Certified) | Power BI
AI/ML:	Regression, Classification, Clustering, Forecasting, Dimensionality Reduction, NLP, Computer Vision
Big Data Tools:	Apache Airflow, AWS: S3, EC2, Amazon Redshift
Dev-ops Tools:	Git, Docker, MLflow, DVC, Linux/UNIX