



Data science master's student with over 2.5 years of industry experience analyzing data and providing recommendations. Skilled in machine learning, statistics, data visualization, and generative AI. Highly dedicated problem solver, goal-oriented, and an efficient team player. Self-driven and fast learner, looking for full-time opportunities.

## EDUCATION:

**Master of Science in Engineering Data Science - GPA: 3.95**

**Jan 2023 – May 2024**

University of Houston, Cullen School of Engineering – Houston, TX

**Bachelor of Technology in Civil Engineering**

**Aug 2015 – May 2019**

Manipal Institute of Technology – Manipal, India

## TECHNICAL SKILLS:

- **Programming:** PyTorch | Python (advanced) | HTML, CSS (basics)
- **Database:** MySQL Redshift, Snowflake | Vector Database (FAISS, Chroma) | Graph Database (Neo4j, NebulaGraph)
- **Analytical Tools:** Tableau, Power BI, Microsoft Excel (advanced), Microsoft PowerPoint
- **AI/ML:** Regression, KNN, ANN, Decision Trees, Random Forest, SVM, Apriori, Dimensionality Reduction (PCA, SOM, t-SNE), NLP, RNN, LSTM, Transformers, Explainable AI (XAI), Large Language Models, CNN, Object detection, image segmentation, pose detection, Prompt Engineering, Retrieval Augmented Generation, Knowledge Graph
- **Dev-ops Tools:** Git (GitHub, Bitbucket), Docker, Kubernetes

## WORK EXPERIENCE:

**Generative AI - Research Assistant - AIceberg | Houston, TX**

**June 2023 – Present**

- **Fine-tuned large language models** on a custom dataset using **LoRA** and **QLoRA** fine-tuning methods and quantization of the models.
- **Grounded** the finetuned LLM with external non-parametric knowledge through **RAG** and **Knowledge Graphs** and generate quality response through prompt engineering to tackle **hallucination** in LLM.
- Developing and experimenting with novel techniques in **feature extraction** using vector embeddings and **prompt engineering**.
- Enhanced research and conducted experiments across multiple **frameworks** (LangChain, LLaMaindex), **embedder models** (SOTA embedders), and **vector storage databases** (simple methods to FAISS) to optimize both the speed and quality of processes.

**Business Analyst - Meesho | Bangalore, India**

**Oct 2022 – Jan 2023**

- Produced leads through data mining and measured impact/performance of webinars and 1:1 training through **A/B testing**.
- Provided **business recommendations** to improve supplier engagement.
- Developed and maintained **analytical dashboards** utilized by stakeholders to track the L0 metrics of the Supplier Activation charter.

**Business Analyst - Swiggy | Bangalore, India**

**Apr 2021 - Aug 2022**

- Implemented **A/B testing** and normalizations to measure the impact/performance of in-house products or features on **Chatbot (CRM)** and formulate necessary business recommendations.
- Improved the **CPO by 15%** by changing nomenclature of a bot disposition. Reduced 95th percentile customer wait times during peak hours by **60%** by balancing the load.
- Utilized Power BI to develop and maintain smart, compelling **analytical dashboards** to monitor KPIs, identify trends, and monitor company initiatives and agents' performance.
- Conceptualized the formulation of various metrics (active agents) and enhancement of a bot efficacy metric.
- Conducted **driver analysis** on key metrics to identify potential improvement areas in the Swiggy Chatbot flow.
- Collaborated with enterprise **data warehouse**, data governance, and business teams on **data quality issues**, as well as architecture of **data repositories** or fact tables under my purview.

## TECHNICAL PROJECTS:

**Illuminating LLM Responses (XAI) with GPT 3.5 and T5**

Leveraged **LangChain** framework on **GPT 3.5** to create high-quality **QA downstream task** response on **custom documents**. Fine-tuned the **T5** LLM to generate headlines from responses and employed **contextual similarity** (cosine) on the **embeddings** of T5-generated headline and the document's title to **retrieve** the specific document that was used to generate a response.

**Keyword extraction using Attention mechanism in BERT**

Leveraged **contextual embedding** from the **last hidden layer** of the output of BERT to extract top 5 keywords from any a text body.

**Hotdog or Not Hotdog**

Constructed a custom image classification model that leverages **Inception V3** which could produce an accuracy of 93.67%

**Resume Parser**

Built a **Custom Name Entity Recognition model** using spaCy. Model was trained on a training dataset containing samples of hard skills and soft skills, to extract skills from PDF files.

**HackerEarth: Adopt a buddy Challenge**

Built a **multi-label machine learning model** stacked with **XGBoost** and **Light GBM** algorithms, to determine type and breed of the animal. The model got me a place in top 8% on the leaderboard.

**Content-Based Movie Recommender System**

Built a web app on an end-to-end movie recommendation system using Cosine Similarity, from the data scraped from the TMDb website by calling the API. Wrapped it in a Flask API and deployed it onto Heroku.