MAYANK MUSADDI

Machine Learning Engineer

+91 820 912 5754 | in/musaddi
musaddi97@gmail.com | github.com/mayankmusaddi

ABOUT ME

Machine learning engineer with over 3 years of experience in industry and academia, specializing in research and implementation in the field of Natural Language Processing (NLP).

SKILLS

Languages: Python, C++, Javascript, SQL

Technologies: NLP, GenAI, LLM, AutoML, MLOps, RAG, Responsible AI, Prompt Engineering, Redis, Docker, Kubernetes

EXPERIENCE

June/2022 – Ongoing
Product Engineer — **Sprinklr, Inc.**

- Developed and deployed a **conversational chatbot powered by Large Language Models (LLMs)** on the Sprinklr platform, simplifying user interactions and data visualization across multiple product suites, including customer care and social media insights.
- Spearheaded the **research and development of an entity linking service**, optimizing CPU and GPU utilization, along with Redis calls, to handle high operations (about 100 ops) at minimal cost. This service continuously analyzes live Twitter feeds, providing brands with real-time insights linked to Wikipedia entries.
- Engineered a **live insights service for Amazon sellers**, leveraging social media data to offer dynamic product analysis.
- Contributed to the development and maintenance of Kratos, an **AutoML infrastructure capable of parallel training and deployment** of multiple machine learning models in a scalable and cost-efficient manner, utilizing optimization techniques such as **ONNX** and resource auto-tuning.
- Managed Smartintents, a service responsible for intent detection in customer care conversations. Achieved a **20% cost reduction** through improved **Kubernetes autoscaling, model quantization, and ONNX optimization**, while ensuring target stability.

May/2019 – May/2022
Undergraduate Research Assistant — **CCNSB Lab, IIIT-H**

- Worked as a research student under **Professor Nita Parekh** in the domain of Bioinformatics to use **deep learning approaches to model high throughput data** as a network to gain insights. (link to bio)
- Published a detailed research in the analysis and understanding of phenotypes linked to stress response in rice thus leading to novel insights in multi-stress tolerant crops **(Sanchari et al., 2022)**.

June/2021 – July/2021
Machine Learning Intern — **Sprinklr, Inc.**

- Created a pipeline for **knowledge distillation** of large models for language modelling task. Used it to distill large RoBERTa and XLM-RoBERTa models fine-tuned on Twitter specific data, to smaller pre-trained distilRoBERTa model which is half the size and twice as fast.
- Created a web application to **host and visualize models and datasets** that is used by the Sprinklr AI team, using Python Streamlit.

May/2019 – July/2019
Machine Learning Intern — **VISIRIS Innovation(P) Ltd.**

- Automated data harvesting from live Football Match Telecast using **Python OpenCV, MATLAB and Flask**. Data like player positions, team classification, types of passes and the location of ball and goal post was fetched from the match images using concepts of **Morphology and Homography**.

EDUCATION

April/2021 – May/2022
MS by Research — **International Institute of Technology, Hyderabad**
Published journal paper and defended thesis on interdisciplinary research in Bioinformatics and Machine Learning under the supervision of Professor Nita Parekh.
Teaching Assistant: Computer Programming, Software Architecture, Distributed Systems and Operating System.

April/2017 – Mar/2021
B.Tech in Computer Science — **International Institute of Technology, Hyderabad**
GPA: **9.39/10** | Major GPA: **9.58/10**
Dean's Merit List Holder: Excellence in academics for all semesters (**Top 5% in batch**)
Elected member of Student's Parliament representing a batch of 200 students
Design and App Team head for Annual Fest for IIIT-H 2018-2020

PROJECTS

LLM Chatbot
Persona based Task Automation
Created a **generic infrastructure for LLM powered chatbots** that could interact with the user, learn their persona and tasks, eventually being able to automate flows through tool use and knowledge base retrievals. (Patent Pending)

GNNs
Spatial Transcriptomics Embeddings
A **hyperbolic graph neural network** approach to identify hierarchies in cell types from spatial transcriptomics data obtained from seqFISH+.

Data Visualization
NetREx — **application link**
Procuring and **refining dense networked data** capturing gene coexpression. Representing the data via an independently developed end to end web application for data visualisation using Sigma JS and Express JS.