Isita Polamarasetti

ip22@rice.edu | (832) 806-9945| Github | Linkedin | PortFolio

Education

Rice University- GPA: 3.8 / 4.0 **08/2023 - 12/2024**
Master's in Computer Science *Houston , TX*

Gitam University - GPA: 9.43 / 10.0 **06/2019 - 04/2023**
Bachelor of Technology in Computer Science and Engineering *Visakhapatnam, India*
- Awarded a **25%** scholarship each year for academic excellence at Gitam University.
- Top **10%** in **Leetcode**.

Skills

Languages: C++, Python, Java, JavaScript, Typescript, Matlab, R
Web Development: React.js, HTML, CSS, Node.js, jQuery, Bootstrap, REST
Data Science: NumPy, Pandas, Matplotlib, Sklearn, TensorFlow, PyTorch, PySpark, Plotly, NLTK, Spacy, Ml-Flow, Pydantic , Langchain
Databases: SQL, PostgreSQL, NoSQL, Snowflake
Tools/Platforms: GIT, Linux, Pytest, BeautifulSoup, Selenium, Hadoop, Docker, Tableau, CI/CD, ETL, AWS, Excel, Microsoft Tools

Experience

Baylor College of Medicine **08/2024 - Present**
Machine Learning Intern *Houston , TX*

- Fine-tuned foundational models, including "**scGPT**", **"GPT-2"** on Large patient single-cell data. Examined gene activation differences between normal and tumor cells, enhancing predictive accuracy and applicability.
- Explored the **CelltoSent** model to interpret label predictions for gene sequences and scrutinized LLM hallucinations using data from **30k+** cells. Incorporated **KGs** and **RAG** pipelines, improving cell type prediction accuracy to **96%**.
- Leveraged **KDTree** to evaluate ligand-receptor interactions across **50k+** cell pairs in single-cell mouse embryo data, revealing insights into cellular signaling pathways.

Intelligent Medical Objects - IMO Health **05/2024 - 08/2024**
Natural Language Processing(NLP) Engineering Intern *Houston , TX*

- **Rare Disease Prediction:** Analyzed data from **3.5k+** individuals diagnosed with and without PAH (Pulmonary Arterial Hypertension), creating detailed charts to illustrate disease stages and facilitate a clearer understanding of co-occurring conditions and Mutual information. Performed statistical tests to draw meaningful comparisons.
- Enhanced systematic literature reviews using LLMs, reducing time by **41%** and increasing accuracy by **26%**. Optimized extraction and prompt engineering **(COT)** from **4** to **2** steps, added checks,and utilized **Pydantic** and **LLM chain** for multimodal data retrieval.
- Conducted time series analysis on MMR and HPV vaccine trends, identifying peak months at **2.8%** (MMR, 2018) and **2.0%** (HPV, 2016). Highlighted seasonal patterns and milestones to inform public health strategies

Rice University **12/2023 - 08/2024**
Teaching Assistant (COMP 543) | Student Computing Tech lead | Grader (STAT 502) *Houston , TX*

- Collaborated with **Professor Christopher Jermaine** on **'Grad Tools and Models for Data Science,'** conducting **15+** office hours.
- Evaluated coursework for **'Neural Machine Learning 1 (Comp 502)'** with **Erzsebet Merenyi**, managing **15%** of coursework for **300** students and facilitating the mastery of data science tools such as **MySQL, Pyspark, tensorflow** and **AWS**.
- Led a team of **10** at Rice University's IT Help Desk, resolving **80+** technology-related queries weekly, account management, network troubleshooting, and software deployment, documented and escalated issues.

Footsteps LLC (Amazon DSP) **02/2023 - 07/2023**
Software Engineering Intern - Data Automation (Remote) *California, USA*

- Applied statistical modeling and machine learning, including LOF and gradient boosting, to Amazon driver data, cutting delivery times by **11%** and improving fuel efficiency by **13%**. Conducted **A/B testing** to optimize logistics, including driver shift patterns and incentive programs, yielding **$25k** in biweekly bonuses.
- Developed automation solutions for Amazon logistics, integrating APIs and custom scripts, saving **21** hours per week in data entry and reducing errors. Leveraged **OpenStreetMap** API to build a React-based logistics site with dashboards, boosting efficiency by **33%** and enhancing delivery tracking.

Pharmapro Tivra Health LLP **10/2022 - 07/2023**
Data Science Intern (Remote) *Mumbai, India*

- Implemented **NLP classifiers** and matchers for complex biomedical clinical trial data, achieving a **35**% increase in accuracy through the application of machine and deep learning algorithms like **Token-based Matching** and **Bert**.
- Designed interactive dashboards and data-driven applications using **Tableau** and Python libraries to visualize medical specialties, highlighting the distribution of specialized doctors, including experience, demographics, and other factors.
- Orchestrated integration of healthcare APIs from industry-leading sources like **nih.gov.in** and **credihealth.com** to automate extraction of structured medical data, resulting in a **85%** improvement in data acquisition speed and accuracy.

Projects

- **Efficient Hashing Using Huffman Coding:** Integrated Huffman coding to enhance data manipulation and optimize space utilization by **47%**. Demonstrated proficiency in advancing techniques that reduced memory usage in HashMaps.
- **PixPlate:** An innovative GenAI app offering personalized recipes based on dietary needs. Users can select products or upload photos of items in the fridge. The app utilizes **TensorFlow** and **LLMs** for object detection and provides **5** tailored recipe recommendations, leveraging **AutoVAE** to adapt to user preferences. Built with **React**, **MongoDB**, **Bootstrap**, and deployed on **AWS**.