Saksham Gupta

gupta.saksham98@gmail.com ❖ (317) 717-0962 ❖ Linkedin ❖ Github ❖ Blog ❖ Indianapolis, Indiana

WORK EXPERIENCE

Biogen **June 2023 – August 2023**
Pharma Manufacturing Data Science Intern *Durham, North Carolina*

- Partnered with Manufacturing groups, utilizing machine learning and NLP techniques to optimize manufacturing operations.
- Engineered robust and consistent metrics using machine learning, particularly NLP, empowering leadership to gain actionable insights into workflow patterns, fostering productivity and operational improvements.

Indiana University School of Medicine **September 2022 – December 2023**
Research Assistant *Indianapolis, Indiana*

- Engineered LLM (Large Language Model) (BLOOMZ, LLAMA, T5) based NLP pipeline to extract adverse event data from clinical trial notes.
- Designed and 3D-printed lab containers using generative AI and constraint optimization tools.

ZS Associates **August 2020 – August 2022**
Data Science Associate Consultant *Bangalore, India*

- Led a cross-functional team to launch a Plotly-Dash-based NLP web application, streamlining organization-wide data visualization and integrating high-capacity data pipelines and APIs for unified data access.
- Directed efforts in propensity modeling, market segmentation, and linear optimization for a leading US bio-pharma client; transitioned legacy R workflows to PySpark, resulting in 150% speed improvement, and expanded project scope to cater to three varied disease portfolios.
- Implemented a nationwide Transportation optimization proof-of-concept for a premier Japanese pharma client, creating an interactive Tableau Dashboard for transport scenario simulation, and leveraging Geo-spatial algorithms for optimal treatment center identification.
- Co-hosted a corporate podcast highlighting the experiences of leading data scientists at ZS and investigated the potential of quantum computing techniques in accelerating core life sciences tasks, as part of an internal company initiative.

Stylumia Intelligence **February 2020 – July 2020**
Product Engineering Intern *Bangalore, India*

- Applied Deep Learning for segmentation and attribute extraction from fashion image datasets and aided in developing a data labeling tool.
- Unified disparate image datasets into COCO format using OpenCV, pycocotools, and Numpy.
- Boosted Intersection-Over-Union (IoU) of semantic segmentation models by over 50%, enhancing model efficiency and quality of attributes.

ZS Associates **May 2019 – July 2019**
Data Science Associate Intern *Bangalore, India*

- Utilized storytelling, Explainable AI (XAI) techniques, and advanced data visualization to enhance interpretability of complex ML algorithms.

American Express **May 2018 – July 2018**
Technology Intern *Gurgaon, India*

- Constructed Deep Learning-based financial forecasting tool, integrating time-series ensemble methodologies to enhance forecasting accuracy.

EDUCATION

Indiana University - Purdue University, Indianapolis **December 2023**
Masters of Applied Data Science *GPA: 3.96/4*

- Coursework: Biostatistics, Computer Vision, Quantum Computing, Remote Sensing, Database design, Cloud Computing etc.
- Teaching Assistant: 2D Animation, Advanced 2D Animation, Motion Graphics

Vellore Institute of Technology **September 2020**
Bachelors of Technology, Computer Science and Engineering *GPA: 3.8/4*

- President - Deccan Chronicles (The Literary Society) / Vice President - Bulls and Bears (The Finance Club) / Core Committee Member - IEEE Computer Society and IETE India / Radio Show Host - VIT Community Radio

SKILLS

- **Programming**: Python, R, Julia, PySpark, Git, LATEX, Matlab, MarkDown, HTML, Javascript
- **ML Packages**: Numpy, Pandas, Scikit-learn, Tensorflow, Pytorch, JAX, Huggingface, Langchain
- **Visualization**: Tableau, Plotly, Matplotlib, Seaborn, D3.js, Power BI, Alteryx
- **Other**: ArcGIS, Airflow, Google Earth Engine, Github, Microsoft Office, MLFlow, Sagemaker, Kedro, BentoML, Dask
- **Communication**: English (native), Hindi (native), German (intermediate), Russian (beginner)

RESEARCH EXPERIENCE / PROJECTS

- Using Large Multi-Modal (Text + Images) models to generate storylines and artwork for graphic novels - Involves digging deep into model embedding spaces, ControlNet based techniques for more accurate image generation, and automated prompt engineering.
- Using LangChain and Vector storage tools like Chroma/Pinecone to generate clinical trial embeddings to optimize trial search.
- Paper on LLM prompt injection, possible mitigation methods and public policy-based AI regulations required coming soon!
- Designing my own esoteric programming language, based on ASCII art. Because why not. Also writing my own LISP compiler.
- Participated in the International Workshop on Pattern Analysis and Applications, Indian Statistical Institute, Kolkata.
- Paper on Explainable ML accepted at the Frontiers of Intelligent Computing: Theory and Applications conference, NIT Surathkal.
- Worked with Prof. Santhi V (CS HOD) on SAR (Synthetic-Aperture Radar) Image de-noising to remove salt and pepper noise from satellite based remote sensing systems for the Indian Space Research Organization.