Area of Interest & Scholastic Achievement

- **Statistics | Machine Learning | Deep Learning | NLP | Generative AI**.
- Achieved **3rd place in** the AI/ML GC at IIT Bombay.

M.Tech Thesis & Industrial project

Multi-modal RAG based LLM for Information Retrieval | *Guide:* ***Prof. Asim Tewri*** *(May'24-Present)*

Objective: Develop RAG based large language model for retrieving the information from unstructured data.

- Innovated a **RAG** application with integrated feedback for improved answer retrieval, utilizing **LanceDB** for vector storage, **Docker** for containerization and deployed on **AWS**.
- Enhanced context fetching **accuracy** by **9%** through the implementation of **reranking methods** and **AI agents** using **Langchain** for query processing, with answers generated by **Mistral AI**.
- Developed a **Textual-Visual** system enabling efficient handling of text and image inputs or outputs using **PyTorch and Pillow**, achieving **92%** answer relevancy using **RAGAS** (benchmarked against ChatGPT-4).
- **Impact:** Generate the correct answers (with associate text and images) from large **unstructured data** source (PDFs) and collect and utilize **user feedback** for customized answer.

Generative AI oil analysis report generator | *Guide:****Prof. Asim Tewari*** | ***Industrial Project*** *(May'24-Aug'24).*

- Developed and deployed a **web application** for **Spectrometric Oil Analysis**, using **HTML**, **JavaScript**, **PHP**, and **Flask APIs.** Application evaluates the health of **6 types** of equipment by analyzing **24+ elements** in oil.
- Engineered **Machine Learning Recommendation System** with real data of **21000** oil samples across **24** types.
- Achieved **96.82% accuracy** in predicting oil types using multi-class classification model with **Random Forest** algorithm and 5-fold **cross-validation.** Further evaluations conducted using **SVM, k-NN and Gradient Boosting**.

Competition & Seminar

Cancer Lesion Prediction for Long-Tailed Distribution (AI-ML GC) | *IIT Bombay* *(Jan'24-May'24)*

- Fine-tuned and experimented with **ConvNet, EfficientNet and VGG16** to determine top-performing model.
- Experimented with **Focal Loss, ASL & CE** loss functions to tackle multi-class data with an imbalance ratio of 58.

A review on the Digital Image correlation | *Guide:****Prof. Asim Tewari*** *(Jan'24-May'24)*

- Leveraged the μ**-DIC** library for strain analysis on real-time data collected from the **AMTF** lab, comparing the results with VIC software.presented the origins and advancements of **Digital Image Correlation** technology.

COURSE PROJECTS

Finding intersection of two surfaces and finding the equation of curves | *Prof.* ***Shyamprasad Karagadde****(Jan-May'23)*

- Applied **Steepest Gradient descent** and **Kernel Ridge Regression** to find equation of intersection of the curves.
- Developed a **MLP** with custom **activation function** in **TensorFlow** to reproduce the results using Deep Learning.

Mice protein expression dataset for Down's syndrome treatment | *Prof.* ***Amit Sethi*** *(Jan'23-May'23)*

- Conducted **EDA** and feature importance using **Recursive Feature Elimination** to eliminate correlated variables.
- Performed **5-fold cross-validation** with **hyperparameter** tuning on **Elastic Net Logistic Regression, Neural Networks, SVM**, and **Random Forest**, achieving **85% accuracy** with **Linear SVM**.

NLP for Scientific Data Collection from Literature | *Prof.* ***Alankar Alankar*** *(Jan'23-May'23)*

- Developed a **novel approach** to extract **5+** properties of **8+** metals and alloys from Engineering Literature.
- Trained and deployed **NER model** on **custom-annotated** dataset with validation accuracy above **95%+**.

Position of Responsibility

- **Led** the grading system for the MS101 course, overseeing 16 Teaching Assistants and managing the academic progress of 720 students throughout the entire semester. *(Jan'23-May'23)*
- **Laboratory Research Assistant in Makers Space course (MS101).** *(July'23-Dec'23)*
- Managed 120 students for the MS101 course with 12 Teaching Assistants, coordinating quiz preparation and online Moodle administration, enhancing the efficiency of course delivery and assessment and utilized tool.

Key Courses & Technical Proficiency

- Programming Softwares/Tools: Python, SQL, HTML, JavaScript, Streamlit, C,C++, LATEX, Docker, AWS, PowerBI.
- ML/Python Libraries: TensorFlow, PyTorch, Scikit-Learn, OpenCV, Pandas, NumPy, Seaborn.
- Hobbies: Playing Cricket, Listening to Music, Watching Movies.