## AREA OF INTEREST & SCHOLASTIC ACHIEVEMENT

- **Statistics | Machine Learning | Deep Learning | NLP | Generative AI**.
- Secured **3**rd position in the institute-wide AI/ML General Championship (Hackathon) held over 10 days

## M.TECH THESIS

**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.

## INDUSTRIAL PROJECT

**Generative AI oil analysis report generator** | *Guide : Prof. Asim Tewari* *(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**.

## DEEP LEARNING PROJECT

**Visual Multiple Instances Zeroshot Transfer for Histopathology Image** | *Guide : Prof. Balamurgan* *(Jul'23-Dec'23)*

- Utilized **Genomic Data Commons (GDC) APIs** to extract and manage **115 GB** of large datasets.
- Developed model with **ReLU**, optimizers **(Adam, SGD)**, and **MaxPooling**, achieving **86.90%** validation accuracy.

**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.

## MACHINE LEARNING AND STATISTICS PROJECT

**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** | Guide : *Prof. Amit Sethi* *(Jan'23-May'23)*

- Conducted **EDA** and developed a pre-processing pipeline with multivariate feature imputation on **75+ proteins**.
- Identified optimal **hyperparameter** settings in **4 ML models** using **5-fold CV** to determine the best models.
- Performed **Feature Importance RFECV** to reduce data dimension by **50+%** achieving **91%** validation accuracy.

**Maximizing Revenue for Drivers** | *self project* *(Jan'23-May'23)*

- **EDA** and **feature engineering** were utilized to detail fare and payment types, revealing critical patterns in the data. A T-test was executed to measure payment type's impact on fares. Significant variations were confirmed.

## POSITION OF RESPONSIBILITY

**Lab Research Assistant | MS101 Course** *(Jan'23-May'23)*

- **Led MS101 grading**, overseeing 16 TAs and managing academic progress for 720 students throughout semester.

**Student Companion | Institute Student Companion Program, IIT Bombay** *(July'23-May'24)*

- Worked in a team of **235+** coordinators, ensuring a smooth transition of incoming first-year PG students .

## 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, NLTK, spaCY.
- Hobbies: Playing Cricket, Listening to Music, Watching Movies.