Name: Harsh Sagar| Roll No: BS19B013
Indian Institute of Technology Madras
Github Linkedin

EDUCATION

Program	Institution	CGPA/%	Year of Completion
Dual Degree in Biological Sciences	Indian Institute of Technology, Madras	7.94	2019-2024
XII (GSEB)	Best High School, Ahmedabad	80%	2019
X (GSEB)	Best High School, Ahmedabad	87.17%	2017

SCHOLASTIC ACHIEVEMENTS

- **Only** student from **2019 batch** to achieve **Panasonic Scholarship Program**
- **Gold Medal** for Class Rank 1, cleared 1st Level and secured **Int. Rank 3359** in 2nd Level of National Science Olympiad in Class X
- Secured International **Rank 859** in International Science Olympiad in **Class IX** conducted by **Science Olympiad Foundation (SOF)**

RELEVANT COURSES AND SKILL

*Completed Prof. Course, ** Coursera, ***NPTEL

• Data Structures & Algorithms for Biology*	• Biostatistics*	• Fundamentals of Operation Research*
• Statistics for Data Science with Python**	• Convolutional Neural Networks**	• Introduction to DL & NNs with Keras **
• Introduction to Machine Learning***	**Programming Languages:** C, basic SQL, Python, React, React Native	**Hackathons** Univ.Ai

WORK EXPERIENCE

Data Science Intern at FN MathLogic Consulting Services, Gurgaon *May'23 – July '23*

Developing a Conversational Question Answering System using Large Language Models (LLMs)

- Explored **LLM memory retention** via various **model finetuning methods** and **Langchain** framework on **ICICI Lombard Policy QnA data**
- Implemented **Transfer Learning** on Microsoft's **DialoGPT** model which was able to retain chat memory for **4-5 conversations**
- Achieved **60% ROUGE SCORE** on DialoGPT also used **Reinforcement Learning** from **Human Feedback** method with **GPT** gave **25% rouge scor**

Document Parsing for Question Answering using LLMs with Langchain framework

- Transformed **ICICI Lombard policy docs** into chunks, each were converted to embedding from **LLM** or **encoder-only** model's embedding
- Used **Langchain** to retrieve **top k** similar document chunks to user's query. These are given with **prompt** to LLM for answer generation
- Utilized **Flan-T5 model** as **LLM** and **embeddings** from **encoder-only** model of **HuggingFace** platform led to **30% rouge score**

Machine Learning Intern at Street Style Store Company, Delhi *May'22 – Aug'22*

- Worked on **Classification** of the conversation between a user and agent about progress of online order to improve user experience
- Utilized **N-grams** stored as **Bag of Words** & compared **Recall values** for SVM model that achieved a **96.4%** accuracy on the test dataset

Data Science Intern at Supratech Lab, Gujarat Nov'21 – Dec'21

- Identified crucial biomarker genes for IVF that enabled successful implantation, enhancing the effectiveness of IVF treatments
- Employed **PCA** & **t-SNE** for dimensionality reduction & conducted **T-test** calculating **P values** that reduced **57k** GeneIDs features to **6k**
- Used **Feature Importance** function of **Random forest** on data and trained **XGBoost** model that achieved **82% test accuracy**

RESEARCH EXPERIENCE

ML Research Intern
IIT Varanasi
Research Paper
June'21 – July'21

Machine Learning for Milk Foam Analysis, Guide: Prof. Abhishek Dhoble

- Worked on Classification of **Surfactants** to study milk foam quality as demand for **Cappuccino foams** has increased
- **Randomforest** achieved **0.955 roc-auc score** & **88.1% test accuracy**, also applied algorithms eg **SVM, XGBoost etc**
- Explored **Casein** reaction with **surfactants**, this study also got published as the **research paper** in **Springer Nature**

DDP PROJECT
IIT Madras
July'23- Ongoing

Evaluating ML models for Chest X-ray of Diseases, Guide: Prof. Ganapathy Krishnamurthi

- Applied **Transfer Learning** on **Ensemble** of **CNN** models the chest x-ray images of Pneumonia and Normal person
- Implemented Image Data Generator for increasing training data to avoid **overfitting** & obtained **93.7%** accuracy

COURSEWORK

EddyNet: For Pixel-Wise Classification of Oceanic Eddies, [OE5015: Machine Learning for Ocean Engineers]

- Classified sea surface height maps using EddyNet, comprising convolutional encoder-decoder **U-Net** and a **pixelwise classification layer**
- For multiclass classification used **one-vs-all soft dice loss**. Accuracy from **Dice Loss** is **89.08%** and **Categorical Cross Entropy** gave **90.61**

Monte Carlo Simulation [BT 2042: Fundamentals of Biophysical Chemistry]

- Observed polymer assemblies using **Monte Carlo Metropolis criterion** in **matlab** at different interaction energies between atoms
- When interaction energy is **low** the polymer formed aggregate. When only some atoms can interact then polymer had more energy

PROJECTS

Stocks & Crypto Currencies Price Prediction Using LSTM

- Developed Long Short-Term Memory (**LSTM**) models with Keras to predict **Closing price** values based on over a decade of trade data
- While training, past **19** days values used as input to predict next day's value. Achieved mean squared error (**MSE**) of **0.465** on testdata

Netflix Movies and TV Shows Recommendations

- Developed Recommender System using **Content based** method for Netflix movies and TV shows recommendations
- Also provided recommendations for books using **LightFM** hybrid recommender which incorporates both item and user metadata

Quora Question Pairs

- Examined whether the questions in each pair are similar or not by calculating **cosine similarity** between the questions
- Used **GloVe embeddings, tf-idf and doc2vec vectorizer** that **achieved** maximum accuracy of **66%** in latter two methods

POSITION OF RESPONSIBILITY

Computer Vision & Intelligence Club Project Member, Shaastra 2022 *June'21 – April'22*

- Worked on **YOLO v5 model** for detecting circuit components and the **mAP** score of the model evaluated with 25 images is **92.7%**
- Detected **terminal points**, **nodes** in circuit using **BFS algorithm** and generated netlist about connectivity of components with others

Biogen Super Coordinator, Shaastra 2022 *July'21 – Jan'22*

- **Supervised** all the coordinators in **theme ideation** & related events to be organized related to biotechnology under the Biogen Team
- **Collaborated** with other teams of Shaastra which help organizing event and also assist them with the **publicity** of the event

Public Relations Volunteer, Shaastra 2020 *Sept'19 – Feb'20*

- Reached out to various **social media** accounts as a part of PR activity for our social campaign named **"BLINK ".**
- Planned and executed **"Walk in the Dark"** and various activities for our stall at KV ground and spread **campaign motto**

EXTRA-CURRICULAR ACTIVITIES

Sports	• Selected for NSO Fitness program and also participated in Samanvay Marathon 2019