**Harsh Sagar**

Indian Institute of Technology Madras +91 6352010302 | [harsh.sagar2107@gmail.com](mailto:harsh.sagar2107@gmail.com) [**Github**](https://github.com/harshs21/Machine-Learning-Projects.git) **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, SQL, Python, React, React Native **Hackathons** Univ.Ai | | | |
| **WORK EXPERIENCE** | | | | | | |
| **ML Engineer Intern**  **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 **finetuning methods** **LoRA** & **QLoRA** on **ICICI Lombard Policy QnA data** * Implemented **Transfer Learning** on Microsoft’s **DialoGPT** modelthat retained 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 score** * **Nodejs** based frontend interacted with user. **Dialogflow** was used to detect intentions from user messages * **FastAPI** based backend was used to call **LLM model** to generate answer and send them back to frontend | | | | | |
| *Document Parsing for Question Answering using LLMs with Langchain framework* | | | | | |
| * Transformed **ICICI Lombard docs** into chunks, each were converted to embedding using **LLM/encoder-only** model * Used **Langchain** to retrieve **top** similar chunks to user’s query which are given to LLM for answer generation * Utilized **Flan-T5 model** as **LLM** & **embedding** of **HuggingFace** platform led to **30% rouge score** on test dataset | | | | | |
| **ML Engineer Intern**  **Street Style Store**  *May’22 – Aug’22* | *Conversation Classification for Enhancing User Experience* | | | | | |
| * Worked on **Classification** of conversation between user & agent regarding online order to improve user experience * Utilized **N-grams** stored as **Bag of Words** & compared **Recall values** for SVM model that gave **96.4%** accuracy | | | | | |
| **Data Science Intern**  **Supratech Lab**  **Gujarat**  Nov*’21 – Dec’21* | *Machine Learning Modeling for Optimizing IVF process* | | | | | |
| * Identified crucial biomarker genes for IVF that enabled successful implantation, enhancing **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** of **Random forest** on data and trained **XGBoost** model that achieved **82% accuracy** | | | | | |
| **RESEARCH EXPERIENCE** | | | | | | |
| **ML Research Intern**  **IIT Varanasi**  [**Research Paper**](https://link.springer.com/article/10.1007/s12161-022-02379-z)  *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 & **Self Supervised** methods for Pneumonia detection * 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** | | | | | | |
| **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 | | | | | | |
| **EXTRA-CURRICULAR ACTIVITIES** | | | | | | |
| **Sports** | * Selected for NSO Fitness program and also participated in Samanvay Marathon 2019 | | | | | |