Harsh Sagar | BS19B013

**Indian Institute of Technology Madras** [**Github**](https://github.com/harshs21/Machine-Learning-Projects.git) **Linkedin**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **EDUCATION** | | | | | |
| **Program** | **Institute** | | **CGPA/%** | | **Year** |
| Dual Degree in Biological Sciences | Indian Institute of Technology, Madras | | [7.94](https://drive.google.com/file/d/1wiZ1pncpspK_zck90slDH5ftgD8gIXZt/view?usp=sharing) | | 2019-2024 |
| XII (GSEB) | Best High School, Ahmedabad | | [80%](https://drive.google.com/file/d/1xmrdVS2298oWn73iEo1GaN1eJgYp_ibj/view?usp=sharing) | | 2019 |
| X (GSEB) | Best High School, Ahmedabad | | [87.17%](https://drive.google.com/file/d/1pOFaTDUWvLbC4xc4l8xxWafx2Zj8UbUX/view?usp=sharing) | | 2017 |
| [**SCHOLASTIC ACHIEVEMENTS**](https://drive.google.com/drive/folders/1C5rvi4_Rzun_o0NMIfEeJDHhl4c2EQt9?usp=sharing) | | | | | |
| * **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**](https://drive.google.com/drive/folders/1p4x1iPdxFF28BiP__tkCy9kdz_vA45Z7?usp=drive_link)\*Completed Prof. Course, \*\*Ongoing Prof. Course, \*\*\*Coursera, \*\*\*\*NPTEL | | | | | |
| * Environmental Monitoring & Data Analysis\*\* | | * Machine Learning for Ocean Engineers\* | * Data Structures & Algorithms for Biology\* | | |
| * Analysis & Interpretation of Biological Data\* | | * Fundamentals of Operation Research\* | * Biostatistics\* | | |
| * Series and Matrices\* | | * Functions of Several Variables\* | * Introduction to Machine Learning \*\*\*\* | | |
| * Principle of Economics\* | | * Computational Economics\* | * Climate Economics\* | | |
| * Statistics for Data Science with Python\*\*\* | | * Convolutional Neural Networks\*\*\* | * Introduction to DL & NNs with Keras \*\*\* | | |
| **Programming Languages** | C, basic SQL, Python, React, React Native | | **Hackathons** | [Univ.Ai](https://drive.google.com/file/d/1ASBXrp73hLvMV7Ea28BP8WnzvfoFIfyZ/view?usp=sharing) | |
| **Libraries** | Numpy, Pandas, Sklearn, Matplotlib, Seaborn, Pytorch, Tensorflow, Keras, OpenCV | | | | |
| [**PROFESSIONAL EXPERIENCE**](https://drive.google.com/drive/folders/11lzEQZD-BcrrKthqAfL6EFLYzOCssnJQ?usp=sharing) | | | | | |
| **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 **ROUGE SCORE** of **60%** while assessing the DialoGPT model on Test dataset for answering the user's queries related to policy * Also used **Reinforcement Learning** from **Human Feedback** (**RLHF**) finetuning method with **GPT2** model that obtained **25% rouge score** | | | | | |
| *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 – August’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** | | | | | |
| **Machine Learning Research Intern, IIT Varanasi, Research Paper (DOI)** [10.1007/s12161-022-02379-z](http://dx.doi.org/10.1007/s12161-022-02379-z) *June’21 – July’21* | | | | | |
| * Worked on Classification of temperature and surfactants to study milk foam quality as the demand for **Cappuccino foams** has increased * **Randomforest** achieved **0.955 roc-auc score** & **88.1% test accuracy**, also used other algorithms such as **Neural Network, XGBoost etc** * Explored several ML algorithms for predictive modelling | | | | | |
| [**COURSEWORK / ACADEMIC PROJECTS**](https://drive.google.com/drive/folders/12g95Mt-vA3lowAEL1m7cPxzHAw7pKHHO?usp=sharing) | | | | | |
| **Evaluating ML models for Chest X-ray of Diseases, Guide: Prof. Ganapathy Krishnamurthi [DDP]** *July’23- Ongoing* | | | | | |
| * Applied **Transfer Learning** on **Ensemble** of **3-4 CNN** models the chest x-ray images of Covid, Pneumonia and Normal persons * Implemented Image Data Generator to avoid **overfitting** by increasing training data and obtained accuracy about **93.7%** | | | | | |
| **EddyNet: For Pixel-Wise Classification of Oceanic Eddies, [OE5015: Machine Learning for OceanEngineers]** | | | | | |
| * Classified sea surface height maps using EddyNet, comprising convolutional encoder-decoder **U-Net** and a **pixelwise classification layer** * Also studied the usage of Scaled Exponential Linear Units (**SELU**) instead of the traditional ReLU + Batch Normalization (**R+BN**) * 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 employing **Metropolis criterion of Monte Carlo** in **matlab** when two atoms of seperate molecule interact * Found assemblies for different **interaction energies** provided when atoms interact with each other separated by a **single lattice length.** * When interaction energy is **low** the polymer formed aggregate. When only some atoms can interact then polymer had more energy | | | | | |
| **Finance Plus, Finance & Economics Club IIT Guwahati** | | | | | |
| * Completed assignments based on concept of **asset** evaluation like future cash flow from lease, rental payments & capital appreciation | | | | | |

|  |  |  |
| --- | --- | --- |
| * Applied statistical concepts to real-world situations, such as estimating **population variance** for quantifying investment risk | | |
| **Real Estate Auction System, [HS4007: Computational Economics]** | | |
| * Created payoff matrix of hawk-dove game to simulate contest for maximizing player payoff, by applying game theory in real estate trade * In other approach, bid value of bidders are assigned using **gaussian distribution** in every round or **compounding** value of previous round * Compared the above models based on difference in final price from the highest private valuation of bidder in minimum number of roun | | |
| **Forecasting CO2 Emission and Evaluating its Impacts on Climate Change, [HS4870: Climate Economics]** | | |
| * Analysed adaptation benefits, costs & population growth projection to depict mortality risk due to climate change from **carbon emission** * **Assessed** the overall financial burden posed by each metric ton of **CO2** released by assigning monetary values to repercussions * **Polynomial Regression** was used to forecast per capita annual carbon emissions for the years 2021-2030 using past records since 1960. | | |
| **PROJECTS** | | |
| *Netflix Movies and TV Shows Recommendations* [*Link 1*](https://colab.research.google.com/drive/1UGwnxdDxB_s6IAIDsA9HB27u2fd1DJIT?usp=sharing)[*Link 2*](https://colab.research.google.com/drive/1d94tfn91XYb4mWwoWRKD_WbXr4aby3dy?usp=sharing) | | |
| * 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* [*Link 1*](https://colab.research.google.com/drive/1FwWV7O-bG6vA4zVgaSink6LktPRJi7Fd?usp=sharing)[*Link 2*](https://colab.research.google.com/drive/1CY2m6nobWHYhlZ0G49kRy0zB93LPrtQ0?usp=sharing) | | |
| * 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 | | |
| *Trading Strategy and Backtest on Time Series data* [*Link 1*](https://colab.research.google.com/drive/1B8i5wQfisrVoOcx8pMQhX0nNvTpJJTPi?usp=sharing) | | |
| * Meticulously designed stock trading strategy and backtest on **TATASTEEL** data from 2021. Calculated the **P/L** of entire backtest at end * **Bought** the stock if current closing price (**CP**) exceed both Support 1 (**S1**) & Simple Moving Average (**SMA**) that ensured potential uptren * **Sold** the stock if current **CP** was below **S1** & 20-period **SMA** that enhanced risk management. Fixed **1% Profit Target** & **0.5% Stop Loss** | | |
| *Stocks & Crypto Currencies Price Prediction Using LSTM* [*Link 1*](https://colab.research.google.com/drive/1rKnkihFkE8F5iKpzHzomMpSbT1e5lCcJ?usp=sharing) | | |
| * 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 are used as input to predict next day’s value. Achieved **mean squared error** of **0.465** on test data | | |
| *Trading Strategy using Momentum* [*Link 1*](https://colab.research.google.com/drive/15Odi1eMphCEIJ08el14UWQvV0mxNppTa?usp=sharing) | | |
| * Developed trading strategy based on momentum indicators using the historical **Closing prices (CP)** of stocks for a subset of the **S&P500** * Implemented functions to resample prices, compute **log returns** and generated **trading** **signal** using **Momentum** for **1 month** period * Evaluated the performance of the resulting portfolio through hypothesis testing using **T-test** and calculated **p vlaues** | | |
| *To-Do List Web Application (Angular)* [*Link 1*](https://github.com/harshs21/Angular-To-do) | | |
| * Developed dynamic & user-friendly To-Do List web application using Angular framework, allowing users to efficiently manage time * Implemented features to add, dele, & update tasks, along with the capability to create & manage multiple sub-todos within each task | | |
| *Rent A Car Application using Angular* | | |
| * Built and maintained a **dynamic Single-Page Application (SPA)** using Angular, empowering efficient car rental management * **Implemented admin functionalities** for cars, brands, colors & services, empowering efficient data management | | |
| *Bus Tracking Web Application* [*Link 1*](https://github.com/harshs21/Bus-Tracking) | | |
| * Developed real-time bus tracking web application for campus transport. Devised user interface & backend logic to display live bus location * Utilized GPS data from buses via APIs to provide accurate real-time coordinates. Implemented features for route visualization, bus arrival predictions, and user-friendly interface. | | |
| **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 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](https://drive.google.com/file/d/1LdGRidNS3mzIoBoHTM8-36VqSfyQkXU5/view?usp=sharing) | |